ESSAYS ON ECONOMETRICS

Multivariate Markov Chains

Author:
Bruno Damásio

Supervisor:
João Nicolau

A Thesis submitted in fulfilment of the requirements for the degree of PhD

in

Mathematics Applied to Economics and Management

ISEG, Universidade de Lisboa

Lisboa, September 2018

ESSAYS ON ECONOMETRICS

Multivariate Markov Chains

Author:
Bruno Damásio

Supervisor:
João Nicolau

A Thesis submitted in fulfilment of the requirements for the degree of PhD

in

Mathematics Applied to Economics and Management

ISEG, Universidade de Lisboa

Committee Chair:

Nuno João de Oliveira Valério, ISEG, University of Lisbon.

Committee Members:

Nuno Paulo de Sousa Arrobas Crato, ISEG, University of Lisbon.

João Carlos Henriques da Costa Nicolau, ISEG, University of Lisbon.

Manuel Leote Tavares Inglês Esquível, FCT, NOVA University of Lisbon.

Luís Filipe Farias de Sousa Martins, ISCTE-IUL, University Institute of Lisbon.

Paulo João Figueiredo Cabral Teles, FEP, University of Porto.

This work is dedicated to my daughters Luz and Carmen for teaching me the art of the multiplication of love. I also dedicate this Thesis to the *epsilon* that will be born healthy next March.

ABSTRACT

This Dissertation is about Markov chains and their role in economics and in econometrics theory. Four essays on the Markov chain approach are presented.

We start by illustrating the analytical potential of multivariate Markov chains in the field of economic history, in particular with regard to a test of the Schumpeterian hypothesis of creative destruction.

Then, we ilustrate the flexibility of Markov chains, and their pertinence to situations that go beyond their traditional applicability: i) how can a Markov chain play the role of covariates; ii) how can a Markov chain representation be useful to compute expected hitting times.

Finally, we present a new methodology for testing and detecting multiple structural breaks in multivariate Markov chains, where the dates at which the structural breaks occur are unknown.

ACKNOWLEDGEMENTS

The guidance of my supervisor, João Nicolau, throughout this work was incredible: I would like to thank you for the privilege and pleasure of working with you.

Two essays included in my dissertation were made in coauthorship. I would like to express my gratitude to Francisco Louçã and Sandro Mendonça for their contributions to the quality of this work.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\mathbf{E}[X]$	The expected value of X
$P(A)$	The probability of A
$P(A \mid B)$	The conditional probability of A given B
$\hat{P}(A)$	The estimated probability of A or the estimator of $P(A)$
P_{ij}	The probability $P(S_t = j \mid S_{t-1} = i)$
$\mathbf{P}_{\bullet k}$	The is the k-th column of the transition probability matrix $\mathbf{P}$
$\mathbf{P_t}$	The transition probability matrix for the period t
$Var(A)$	The variance of X
$1\{A\}$	The indicator function, equals 1 if A is true and 0 otherwise
$\mathcal{F}_t$	The σ-algebra generated by all events up to time t
$Var(A)$	The variance of A
$\{(X_t); t = 0, 1, 2, ...\}$	The stochastic process X_t
q_l	The $l - th$ percentile
Φ	The cumulative standard normal distribution function
$\mathbf{x}$ (bold lowercase Roman letter)	The column-vector $\mathbf{x}$
$\mathbf{x}'$	The transpose of the column-vector $\mathbf{x}$
$\boldsymbol{\theta}$ (bold lowercase Greek letter)	The column-vector of parameters $\boldsymbol{\theta}$
$\boldsymbol{\theta}^0$	The true value of the column-vector of parameters $\boldsymbol{\theta}$
$n_{i_1 i_0}$	The number of transitions from the state i_0 to the state i_1
T_{x_1}	The hitting time $\min\{t > 0 : y_t \geqslant x_1\}$
$\xrightarrow{p}$	Convergence in probability
$\xrightarrow{d}$	Convergence in distribution or weak convergence
$\mathrm{vec}(A)$	The vectorization of the matrix A
$\otimes$	The Kronecker product
$\boldsymbol{\Pi_j}$	The vector of the stationary probabilities for the j-th segment
$a \equiv b$	a defines b

ACRONYMS

AIC Akaike information criterion

AR(p) Autoregressive Process of order p

DGP Data-Generation Process

DM Diebold and Mariano Test

ECB European Central Bank

ET Expected Hitting Time

ETC Expected Time Curve

FPE Final Prediction Error

GARCH Generalised Autoregressive Conditional Heteroskedasticity Model

HAC Heteroskedasticity and Autocorrelation Robust Standard Error

HOMC Higher Order Markov Chain

HOMMC Higher Order Multivariate Markov Chain

HQ Hannan-Quinn information criterion

LogL Log-Likelihood

LR Likelihood Ratio

MC Markov Chain

INAR Integer-valued Autoregressive Model

MLE Maximum Likelihood Estimator

MMC Multivariate Markov Chain

MTD Mixture Transition Distribution Model

MSE Mean Square Error

OLS Ordinary Least Squares

QMLE Quasi Maximum Likelihood Estimator

r.v. Random Variable

SIC Schwartz information criterion

SE Standard Error

TPM Transition Probability Matrix

VAR Vector Autoregressive Model

1 INTRODUCTION

This Thesis consists of four essays on Markov chains and its potential for furthering economic understanding on a variety of empirical issues and analytical challenges. Over recent decades both linear and non-linear time series models have played a crucial role, not only in Economics but also in related fields such as Finance and the Social Sciences. During the early 1970s the linear Box-Jenkins ARMA methodology enjoyed enormous acceptance. Since then, non-linear models have become increasingly popular, such as threshold model [93, 94], Markov-switching model [40] and (G)ARCH models [15, 30], see also Nicolau [63]. All these models share a common feature: the dependent variable is continuous; notwithstanding, in some circumstances the dependent variable may be categorical. Categorical time series arise frequently in economics and other sciences, e.g. international debt ratings, labour market flows, variables that are characterised by state transitions or path-dependence in general. From another perspective, a discrete reconstruction of a continuous stochastic process may be particularly useful since it allows us to extract some apparently imperceptible patterns from the data. Qualitative variables allow us to model structural changes in the processes, to focus on the frequency and length of certain events (such as bear and bull markets), and to isolate some characteristics of the processes: an interaction between a qualitative variable and a continuous one may point out some patterns, e.g. in business cycles, such as velocities of the transition between states over time. Thus, modelling qualitative events is extremely important in Economics, in order to establish accurate policy guidelines to promote economic growth, employment and to guarantee the sustainability of the welfare state. However, the aforementioned linear and non-linear models designed for continuous variables are not applicable when the dependent variable is categorical. The multivariate dynamic probit models are a class of meaningful models, but these models can only deal with two state processes and Economics often entails situations where we have more than two states. A logical option could be multivariate Markov chain models. However the huge number of parameters, when the dynamic of the model is of higher order and the number of states and series is large, make the estimation impossible due to lack of degrees of freedom. The mixture transition distribution probit model (MTD-Probit) solves this problem by modelling through discrete time Markov chains with arbitrary large domains, [64]. The model has some advantages over the Raftery's MTD [75] due to the absence of constraints, thus facilitating the estimation procedure. Therefore, this methodology allows us to efficiently estimate multivariate discrete processes, and in particular multivariate Markov chains, based on reducing the number of parameters. In general, the Markov chain approach has some advantages over alternative approaches: it is a non-parametric approach and can capture non-linear relationships between variables, unlike INAR family models [54]. Consequently, it calls for an extension of the literature regarding the estimation methods of discrete and categorical variables.

The main objective of this Thesis is to extend the literature concerning the treatment of discrete time series, mainly through the improvement of the Markov chains estimation framework. As such, this Dissertation has three main goals.

First, we demonstrate the relevance of considering multivariate Markov chains, even when the phenomenon under study is of a continuous nature (Chapter 2). We show that Markov chains can be a valuable tool to detect Patterns of causality in the presence of a nonlinear relationship, in the situation where standard VAR models fail to reject the hypothesis of absence of linear Granger causality. An example of the model applied in economic history is presented, more precisely the revolution of two commercial sea carriage technologies in the early 19th century: British steam and sail merchant vessels.

In this Chapter, *Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition*, we find that while a vector autoregression approach would be an obvious choice for modelling structural relationships in multivariate processes, the results might be unsatisfactory, as no trace of Granger causality is found. One reason for these failures to identify interactions may be linked to the presence of nonlinear dynamics. Next, we show that a reconstruction of a former continuous process into a discrete one allows a multivariate Markov chain analysis to check whether density functions, and not only first moments, are time-dependent among variables. This modelling strategy succeeds in picking up directional powers between the historical paths of sail and steam technologies. In particular, we detect that while the insurgent (steam) does not impact the incumbent (sail) some effects are produced from the incumbent (sail) to the insurgent (steam) technology. We argue that this finding makes sense in the light of the existing literature on technological change in maritime history.

Second (Chapters 3 and 4), we demonstrate the potential and the flexibility of a Markov chain model using two different approaches. In fact, usually Markov chains are considered as an end in itself. This means that the main goal of an application with Markov chains is usually the computation of the predicted states probabilities. We showed that we could go further, by, for example, computing expected hitting times (Chapter 3); or use the predicted state probabilities to predict a certain continuous variable in a forecasting problem (Chapter 4).

More specifically, in Chapter 3, *Combining a regression model with a multivariate Markov chain in a forecasting problem*, a new concept is presented: the MMC usage as covariates in a regression model in order to improve the forecast error of a certain dependent variable. We assume causality from the latter to the former. In this context, of an endogenous variable that depends on some time-dependent categorical variables, we argue that a multivariate Markov chain approach, that takes advantage of the information about the past state interactions between the categorical variables, may lead to a substantial forecasting improvement of that endogenous variable, confirmed through Diebold-Mariano tests and a Monte Carlo experiment.

In Chapter 4, *The changing economic regimes and expected time to recover of the peripheral countries under the euro: a nonparametric approach*, we focus on an important concept in stochastic analysis, the expected time to cross a given threshold. Moreover, we apply a new method to compute the expected time to recover from a negative (or positive) shock. The method relies on a Markov chain representation of a continuous variable of interest. We apply this method to assess the impact of the monetary regime change that occurred from 1999 onwards on the dynamics of the economies of Europe's peripheral countries.

Third, Chapter 5, *Time inhomogeneous multivariate Markov chains: detecting and testing multiple structural breaks occurring at unknown dates*, proposes a methodology for testing multiple structural breaks occurring at unknown date intervals in Markov chains, that can also be applied to multivariate Markov chains. This is a topic that has received little attention in the literature, despite its great relevance. The research on discrete time Markov chains often assumes homogeneity of the process in the sense that the stochastic transition probability matrix is assumed to be time invariant. However, ignoring the nonhomogeneous nature of a stochastic process by disregarding the presence of structural breaks can lead to misleading conclusions. Markov chains are not an exception and it is relevant to have access to mechanisms that allow testing for inhomogeneities. Indeed, structural breaks are a common issue in Economic phenomena due to their intrinsic complexity and interdependence between variables. However, despite several studies involving Markov chains, few studies involved the issue of inhomogeneity. Two examples are Tan and Yılmaz [91], who used a likelihood ratio statistic to test one single break in the Markov chain that occurs at a known date and Polansky [72], who presented a method to detect and estimate change-points in Markov chains, although the limiting distribution of the test statistic is unknown so the p-values were computed through bootstrapping. Moreover, the method is restricted to first-order univariate Markov chains. This chapter proposes a new methodology for estimating and testing for several structural breaks occurring at an unknown date time in Markov chains. More precisely, this essay proposes a method to: 1) estimate the break dates; 2) test for structural breaks using nonstandard but known distributions (whereas bootstrap techniques are useless). Our methodology can also be extended to multivariate Markov chains and to higher order Markov chains, provided that the sample size is sufficiently large.

2

MODELLING INSURGENT-INCUMBENT DYNAMICS: VECTOR AUTOREGRESSIONS, MULTIVARIATE MARKOV CHAINS, AND THE NATURE OF TECHNOLOGICAL COMPETITION

Note: This article has been published in co-authorship with Sandro Mendonça in Applied Economics Letters.

At least since Schumpeter, modernisation carries the connotation of "creative destruction". The superior newcomer technology makes the old one redundant. In economic history technological competition has been epitomised by the replacement of sail by steam (see, e.g., Craig [23] and Geels [35]). This momentous change in the profile of mercantile marine paved the way to the rise of the west and the triumph of industrial progress.

Most studies, however, have covered the process of transformation of sea-related activity in the late 19th century, when the steamer was already a stand-alone fully viable alternative (see Pollard and Robertson [73], Mohammed and Williamson [60]). Since the classic findings of North [68], through the insights of Harley [43] to the most recent work by Pascali [69], steam navigation has been taken to be a major driving force behind the sharp reduction of transport times and costs that ushered the first era of globalisation.

Less known is the period before machinery and metallurgy reined supreme, when accommodation was the main feature of a maritime world in transition. This paper looks into this earlier period so as to unpack the dynamics of technological co-existence between the two alternatives until the time in which they became clear substitutes (1860s onwards). This empirical work assesses the first five decades of ascendancy of the insurgent, but still experimental technology of steam when sail dominance was overwhelming and yet advancing its performance.

We find that while a vector autoregression approach would be an obvious choice for modelling structural relationships in multivariate processes the results are quite unsatisfactory. Additionally, there is no trace of Granger causality. One reason for these failures to identify interactions may have to do with the presence of non-linear dynamics. To investigate this hypothesis a multivariate Markov chain approach is applied, to check if density functions (not only first moments) are time-dependent between variables. This modelling strategy succeeds in picking up directional powers between the historical paths of sail and steam. In particular, we detect that while the insurgent (steam) does not impact the incumbent (sail) some effects are produced from the incumbent (sail) to the insurgent (steam) technology. However paradoxical, this finding makes sense in the light of the existing literature on technical change in maritime history.

2.1 ARGUMENT AND APPROACH

The industrial revolution at sea is a relatively little explored issue. Clearly, it was a slow process at first. On the one hand, the old technology was more efficient than is usually assumed: sailing ships were pushing ahead in terms of speed and strength comparing with their immediate predecessors [45, 85]. On the other hand, steamships were not competing on established trades and routes as they were handicapped by difficulties of range and efficiency [1], see also Mendonça [56].

Surely both technologies advanced over time; what is less clear is how they influenced each other in this early period in which sail was a dynamic incumbent and steam was still an uncertain insurgent. One way to investigate this issue is by applying conventional time-series techniques that inquire causality and interdependencies between variables, and by taking into consideration a proxy of economically useful ship sophistication (average tonnage). Here we focus on the British merchant sail and steam fleets between 1814 (the earliest datapoint) and 1865 (a cut-off point generally taken to mark the beginning of the end of sail and the end of the beginning of steam; see, e.g. Harley [43]).

2.2 MODELLING THE INTERACTIONS BETWEEN OLD AND NEW TECHNOLOGIES

2.2.1 *Metrics and materials*

In this paper, we start by proposing a vector-autoregression approach to model the evolution of the two technologies between 1814 and 1865. We then try to get extra leverage through a multivariate Markov chain approach. The aim is to characterise the key features of sail-steam dynamic interdependencies.

The dataset is taken from Mitchell [58], a source that makes available for historians long series of economic and technological statistics. Although providing yearly information about number and tonnage for both British-built sail and steamships, this source has remained under-exploited. This study takes the average net tonnage of the sail and steam fleets as a comparative indicator of economically useful technical progress and monitors growth rates over time. As shipbuilding is highly sensitive to the business cycle and to the expansion of trading opportunities British real GDP is taken as a control variable (the Maddison Project is the source here).

It should be pointed out that, given the non-stationary nature of the processes, we considered the log-difference of series (growth rates) of the average tonnage of sail and steam vessels. This is intended to enforce stationarity (both in mean and in variance) as it is confirmed by augmented Dickey-Fuller tests.

2.2.2 *A Vector Autorregression Approach*

In applied econometrics the joint dynamics of variables invites the development of a vector autoregression (VAR) methodology. Since the Sims critique [83] that modelling $K-$dimensional multivariate stochastic process $\{(\mathbf{y}_t), t = 1, 2, 3, \cdots\}$ in the VAR framework has established itself as a standard tool in econometrics.

VAR models explain a multivariate set of endogenous variables uniquely by their own history, exploring the dynamics of the linear interactions between such variables. Therefore, this approach provides a systematic way to capture linear dynamics in multivariate processes. Past shocks to the growth rate in the average tonnage of one type of ship may impact the performance of the other, and/or vice-versa, with years of delay. It may be that Granger-type causality flows from one sort of technology to the other, but not the other way around. In many respects, impulse-response analysis seems an apt perspective through which to conduct causal inference. In order to investigate the dynamics of the relationship between sail and steam in the earlier part of the 19^{th} century, we consider the standard detection and modelling procedures.

Mathematically speaking, a VAR model of order p can be defined as

$$\mathbf{y}_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{\Phi}_j \mathbf{y}_{t-j} + \varepsilon_t \tag{1}$$

where $\mathbf{y}_t = [y_{1t}, \cdots, y_{kt}]'$ is a $K-$dimensional vector of random variables; $\mathbf{c}$ is a fixed $K-$dimensional vector of intercepts controlling for a non-zero mean possibility; $\mathbf{\Phi}_j$ are $K \times K$ coefficient matrices (for $j = 1, \cdots, p$) and ε_t is a $K-$dimensional white noise process such that $E[\varepsilon_t] = 0_k, E[\varepsilon_t \varepsilon_t'] = \Sigma$ (a nonsingular matrix), and, for $\nu \neq s$ $E[\varepsilon_\nu \varepsilon_s'] = 0$

In order to deploy this linear estimation apparatus on the first moments of the time-series we start by establishing the length of the time-lag. Lag length is usually selected using formal statistical criteria like the likelihood ratio (LR), log-likelihood (LogL), Akaike's information criterion (AIC), Schwarz's information criterion (SIC), Hannan-Quinn (HQ), or the Final Prediction Error (FPE). The diagnostic tests point to a restricted model using a minimally-lagged VAR since the LR, FPE, AIC and HQ tests suggest just one time period (1 year) as lag. It is reassuring that many criteria are convergent, as a single criterion is a weak basis from which to judge a model, and that the AIC and the FPE, which are more appropriate when observations are small (60 or less), point in the same direction. LogL and SIC provide conflicting results, but not convergent. Table 1 displays the conventional lag length criteria tests.

Lag	LogL	LR	FPE	AIC	SIC	HQ
0	147.6038	NA	9.87e-06	-3.012579	-2.932443*	-2.980187
1	162.9642	29.44080*	8.64e-06*	-3.145088*	-2.824544	-3.015519*
2	170.9529	14.81242	8.83e-06	-3.124020	-2.563068	-2.897274
3	177.7178	12.12040	9.27e-06	-3.077455	-2.276096	-2.753532
4	185.2677	13.05495	9.58e-06	-3.047243	-2.005477	-2.626144

Table 1: Lag Lenght Criteria

Table 2 reports the Granger linear causality tests. We find that no causality is detected in either direction. That is, no pattern of cross-influence emerges: the null hypothesis of "no Granger causation" is not rejected either for steam being influenced by events in sail in the previous period or vice-versa. This lack of statistical success in picking up the effect the past technological events of one technology on the other may be due to two reasons. First, no connection exists and hence it is not detected. Second, it does exist but is not being modelled correctly.

Finally, Table 3 summarises the VAR estimation results. The development of steamship technology, i.e. the growth is ship size, is highly correlated with itself but nothing else. Results for the other variables are void. Overall, little is learnt from these (linear) exercises. The stage is now set for proposing another (non-linear) probabilistic approach to capture and model a latent structure of interdependencies as a stochastic process.

2.2.3 *Multivariate Markov Chain Methodology*

A Markov chain is a sequence of random variables $S_t, S_{t-1}, ..., S_0$, defined into a countable space state $E = \{1, 2, ..., m\}$, that is characterised by the Markov property that given the present, the future does not depend on the past as follows:

$$P(S_t = k_0 | \mathcal{F}_{t-1}) = P(S_t = k_0 | S_{t-1} = k_1) \tag{2}$$

Where $\mathcal{F}_{t-1}$ is the $\sigma - algebra$ generated by the available information until $t - 1$. The multivariate stochastic process

$$\{(S_{1t}, \cdots, S_{st}); t = 0, 1, 2, ...\} \tag{3}$$

	H_0 : **does not Granger Cause**		
Variable	Steam	Sail	GDP
Steam	–	0.9421	0.0149
Sail	0.5934	–	0.7915
GDP	0.2456	0.4891	–
Joint Wald	0.4732	0.7851	0.0514

p-values are reported

Table 2: Granger Linear Causality Tests

Variable	**Coefficient**	**(Std. Err.)**
Equation 1: Steam		
Steam(-1)	-0.395**	(0.127)
Sail(-1)	0.489	(0.404)
GDP(-1)	0.643	(1.155)
Intercept	3.473	(4.856)
Equation 2: Sail		
Steam(-1)	-0.003	(0.046)
Sail(-1)	-0.081	(0.146)
GDP(-1)	-0.301	(0.418)
Intercept	2.028	(1.756)

Estimates are presented, se's between parentheses.

** denotes statistical significance at the 5% level

Table 3: VAR Model

is said to be a multivariate Markov chain process (MMC) if an only if

$$P\left(S_{jt}=k\middle|\mathcal{F}_{t-1}\right)=P\left(S_{jt}=k\middle|S_{1t-1}=i_1,\cdots,S_{st-1}=i_s\right) \quad (4)$$

Despite its limited usage, this approach configures a substantial advantage with respect to alternative econometric methods; estimating a MMC *tout court* is an impossible task because the total number of independent parameters grows exponentially with the number of categorical series (following $m^s(s-1)$). To address this issue the mixture transition distribution model (MTD) [75] has been proposed. Some improvements to this model have been proposed in the literature, notably by Chen and Lio [20], Ching, Fung, and Ng [21], Ching and Ng [22], Lèbre and Bourguignon [47], Raftery and Tavaré [76], and Zhu and Ching [99].

A salient model is the MTD-Probit model [26, 64]. The quantity $P\left(S_{jt}=i_o\middle|S_{1t-1}=i_1,\cdots,S_{st-1}=i_s\right)$ is taken as a nonlinear combination of bivariate conditional probabilities as follows:

$$\begin{aligned}&P\left(S_{jt}=i_o\middle|S_{1t-1}=i_1,\cdots,S_{st-1}=i_s\right)^{\Phi}\equiv\\&\frac{\Phi\left[\eta_{j0}+\eta_{j1}P\left(S_{jt}=i_o\middle|S_{1t-1}=i_1\right)+\cdots+\right.}{\sum_{k=1}^{m}\Phi\left[\eta_{j0}+\eta_{j1}P\left(S_{jt}=k\middle|S_{1t-1}=i_1\right)+\cdots+\right.}\\&\qquad\qquad\frac{\left.+\eta_{js}P\left(S_{jt}=i_o\middle|S_{st-1}=i_s\right)\right]}{\left.+\eta_{js}P\left(S_{jt}=k\middle|S_{st-1}=i_s\right)\right]}\end{aligned} \quad (5)$$

where

$$\sum_{k=1}^{m}\Phi\left[\eta_{j0}+\eta_{j1}P\left(S_{jt}=k\middle|S_{1t-1}=i_1\right)+\cdots+\right.$$
$$\left.+\eta_{js}P\left(S_{jt}=k\middle|S_{st-1}=i_s\right)\right]$$

is a normalising constant.

The estimation technique is a two-step procedure. The quantities $P_{jk}\left(i_0|i_1\right)$, $k=1,...,s$ are estimated nonparametrically through the consistent estimators $\hat{P}_{jk}\left(i_0|i_1\right)=\frac{n_{i_1i_0}}{\sum_{i_0=1}^{n}n_{i_1i_0}}$ where $n_{i_1i_0}$ represents the number of transitions from $S_{k,t-1}=i_1$ to $S_{jt}=i_0$. The parameters η_{jk} are thereafter estimated using the maximum likelihood method. For the variable S_{jt} the MLE is

$$\log L=\sum_{i_1i_2...i_{i_s}i_0}n_{i_1i_2...i_{i_s}i_0}\log\left(P_j^{\Phi}\left(i_0|i_1,...,i_s\right)\right). \quad (6)$$

It can easily be proved that $\hat{P}_{jk}$ is a consistent estimator of P_{jk} and then it is straightforward to show that $\hat{\eta}_{jk}\xrightarrow{p}\eta_{jk}$.

The parameters η_{jk} represent the weights of the nonlinear combination: the higher the coefficient, in absolute value, the higher the importance of the respective variable $P\left(S_{jt-1}=k\right)$. As the model is estimated through the ML estimator the inference problem is addressed.

This means that the relevance of a specific bivariate probability, that depicts a concrete variable, can be tested from a statistical point of view.

As we will show in the next sub-section, the multivariate Markov chain methodology and the MTD-Probit specification can be used to capture the multivariate relationships and dependences between two technologies. In fact, unlike some traditional parametric econometric techniques, such as vector autorregressions that only capture linear relationships between variables, the purpose of nonlinear methodologies is to capture complex relationships that go beyond the first moment (conditional mean) or even the second moment (conditional variance) as in multivariate GARCH family models. Notice that the absence of parametric assumptions and constraints (the MTD-Probit model is completely free of super-imposed restrictions) underlying the model allows us to capture a wide range of associations between a set of variables that can only be captured using nonparametric approaches.

2.2.4 *Modelling the dynamic relationship between incumbent and insurgent technologies in the early days of steam*

Let y_{1t} and y_{2t} denote respectively the yearly growth rates of the average tonnage (the ratio tonnage/number of ships) of sail and steam. Let also y_{3t} represent the UK gdp annual growth rate. The MMC process was reconstructed according to the following rule:

$$S_{jt} = \begin{cases} 1 \text{ if } & y_{jt} \leqslant q_{j,20} \\ 2 \text{ if } & q_{j,20} < y_{jt} \leqslant q_{j,40} \\ 3 \text{ if } & q_{j,40} < y_{jt} \leqslant q_{j,60} \\ 4 \text{ if } & q_{j,60} < y_{jt} \leqslant q_{j,80} \\ 5 \text{ if } & y_{jt} > q_{j,80} \end{cases}$$

where $q_{j,l}$ represents the $l - th$ percentile of the process y_{jt}, $j = 1, 2, 3$. Regarding the two technologies, the rationale behind this transformation is as follows. Each technology is labeled into five categories or states of innovation accordingly to its development prowess: 1- very slow movement, 2 - slow movement, 3 - standard movement, 4 - fast movement, 5 - very fast movement. The same rationale can be applied to the GDP to economic contraction (state 1 and 2), economic stabilisation (state 3) or economic expansion (states 4 and 5). The main interest here is to analyse the relationships between these two technologies: sail and steam. Information regarding GDP was considered as control and to accommodate the forces that give context to the interdependence pattern that governs the bivariate dynamics under scrutiny. Therefore, $\mathcal{F}_{t-1}$, the $\sigma - algebra$ generated by the available information until period $t-1$ was expanded. For each period the model for the $j-th, j=1,2,3$ category is

$$P\left(S_{jt}=i_o\middle|\, S_{1t-1}=i_1, S_{2t-1}=i_2, S_{3t-1}=i_3\right)^{\Phi} \equiv \frac{\Phi\left[\eta_{j0}+\eta_{j1}P\left(S_{jt}=i_o\middle|\, S_{1t-1}=i_1\right)+\right.}{\sum_{k=1}^{5}\Phi\left[\eta_{j0}+\eta_{j1}P\left(S_{jt}=k\middle|\, S_{1t-1}=i_1\right)+\right.} \frac{\left.\eta_{j2}\,P\left(S_{jt}=i_o\middle|\, S_{2t-1}=i_2\right)+\eta_{j3}P\left(S_{jt}=i_o\middle|\, S_{3t-1}=i_3\right)\right]}{\left.\eta_{j2}\,P\left(S_{jt}=k\middle|\, S_{2t-1}=i_2\right)+\eta_{j3}P\left(S_{jt}=k\middle|\, S_{3t-1}=i_3\right)\right]} \tag{7}$$

Therefore, the space state is $E=\{1,2,\ldots,5\}$, $m=5$ and $s=3$. It should be pointed out the fact that, here, a fully parameterised MMC involves $m^s(s-1)$ independent parameters, circumstance which, in our case, leads to 250 independent parameters which is an untractable problem due to our data span.

The quantities $\eta_{jl}, j=1,2,3; l=1,2,3$ represent the contribution of each past variable for the $j-th$ variable current state. For instance, suppose that we are analysing sail technology. The dependent variable

$$P\left(S_{1t}=i_o|\, S_{1t-1}=i_1, S_{2t-1}=i_2, S_{3t-1}=i_3\right) \tag{8}$$

is a nonlinear function of sail, steam and gdp past states:

$$\eta_{j1}P\left(S_{jt}=i_o\middle|\, S_{1t-1}=i_1\right) \quad +\eta_{j2}P\left(S_{jt}=i_o\middle|\, S_{3t-1}=i_2\right) \quad + \\ \eta_{j3}P\left(S_{jt}=i_o\middle|\, S_{3t-1}=i_3\right) \tag{9}$$

If we fail to reject the null $H_0:\eta_{12}=0$ but we reject $H_0:\eta_{11}=0$ this means that sail does not depend on steam and, moreover, the current power of sail is not determined by steam's power and thus sail is a dominant technology, given that the current performance of sail only depends on its own past performance. The intercepts η_{j0}, although they have no interpretation, are included in the model as they have been shown to improve fit [64, p.1127], so the respective estimates $\hat{\eta}_{j0}$ are reported.

2.3 ESTIMATION RESULTS

This section depicts the estimation results of the equation 7 for the period 1814-1865. Table 4 points out that the estimates $\hat{\eta}_{j1}$ and $\hat{\eta}_{j2}$ measure, respectively, the impact of sail's and steam's past power on the technology j current power.

On the one hand, it can be noticed that the dynamics that governs sail technology is characterised by a dependence on its own past states ($\hat{\eta}_{11}$=6.7794, significant at the 5% significance level, indicating a strong persistent behaviour) and an absence of influence by what has been going on before in steam ($\hat{\eta}_{12}$ not significant at any of the traditional significance levels). On the other hand, steam technology

Equation	$\hat{\eta}_{j0}$ (Intercept)	$\hat{\eta}_{j1}$ (Sail)	$\hat{\eta}_{j2}$ (Steam)	$\hat{\eta}_{j3}$ (GDP)	$Mean$ LL
1 Sail	-4.6688^{***} (1.6422)	6.7794^{**} (3.3522)	7.9404 (4.9772)	5.2054^{**} (2.3467)	−0.0864
2 Steam	-5.7473^{***} (1.8180)	10.2751^{**} (4.7123)	5.8173^{*} (3.1724)	9.3844^{**} (3.9230)	−0.0901

Estimates are presented, se's between parentheses; $Mean$ LL represents the mean of the log-likelihood function.

***,** and * indicates the statistical significance level, respectively, for 1%, 5% and 10%

Table 4: MTD Probit Estimation

is strongly shaped by prior events in sail ($\hat{\eta}_{21}$ is high and significant). This effect happens to be even stronger (both statistically and substantially) than the influence of steam's own past dynamics on itself ($\hat{\eta}_{22}$ is lower and only significant at the 10% level). Both sail and steam dynamics appear to be coordinated with the general economic environment.

Therefore, one may infer an asymmetrical technological relationship. Against heroic or linear representations of innovation, leadership during the rise of the "insurgent" technology was on the side of "incumbent", that is, the vintage solution of sail. Sail's performance does predict steam's, and this circumstance implies a (statistical) dominant influence of the old on the newcomer. The transfer of leadership to steam, in the sense of the impact of the new technology on the old technology, would only occur beyond the period under analysis, and with devastating consequences for sail [23, 56]. At the core of the transformation of transport there was a complex relationship between contending technologies, a switch later amplified by the continuous investment in invention (see Ferreiro and Pollara [32]) and deployment of new infrastructures (see Gray [38]).

2.4 CONCLUSIONS

This paper addresses, analyses and comments the intriguing relationship between sail and steam at the dawn of globalising industrial capitalism. This paper presents evidence that improvements in the incumbent and insurgent technologies appear interrelated. Statistical results suggest that the mix of technologies in the British merchant marine had co-evolutionary characteristics from early on. That a multivariate Markov chain approach brings some fresh and history-friendly insight is testimony to the need for experimenting with new empirical approaches and for keeping the methodological toolbox plural.

Contrary to explanations that would see sail technology reacting to the competing threat posed by steam, we see that technological relations do not simply appear to be zero-sum. Positive, synergic relationships emerge with the arrival of steam to a maritime world dominated by sail. Moreover, the dynamics was not symmetrical. Evidence is somewhat elusive but tentatively points to a major influence from sail to steamship performance (as measured by average carrying capacity). That steam received an indirect payoff from its co-existence with sail resonates with maritime economic history and systemic visions of technical change. These views have emphasised the importance of technological complementarities: the old/incumbent technology, which was in fact quite alive in terms of innovation, re-invigorated the possibilities of the new/insurgent technology (see, e.g. Madureira [51], Mendonça [56], and Rosenberg [78]. Such a stylised fact should be remembered by industrial policy analysts.

3

COMBINING A REGRESSION MODEL WITH A MULTIVARIATE MARKOV CHAIN IN A FORECASTING PROBLEM

3.1 INTRODUCTION

Note: This article has been published in co-authorship with João Nicolau in Statistics and Probability Letters.

Consider a simple regime-switching model

$$y_t = \beta x_t + \delta z_t + u_t \qquad (10)$$

where z_t is a latent dummy variable that evolves over time according to a homogeneous Markov chain (i.e. $P(z_t = i_0 | z_{t-1} = i_1)$, $i_0, i_1 = 0, 1$). This model and further refinements have been extensively studied in the literature (see Hamilton [39]). In some circumstances the z_t variable may be observable, and in this case standard methods of estimation of β and δ apply. However, forecasting y_t may raise some difficulties because z_t (which is assumed to be a random variable) is not observable in the forecasting period (to simplify one assumes that x_t is a dynamic term, e.g. AR(1), or a a simple trend). In this case a probabilistic structure is needed for z_t, for example a Markov chain, as in regime-switching models. In this paper we analyze the forecasting problem when the y_t variable depends on $s > 1$ discrete or categorical variables (observable during the estimation period), whose dependencies are governed by a multivariate Markov chain. This approach is new in the literature and the closest model to ours is perhaps the regime-switching one cited above. However, in contrast to regime-switching models which only deal with univariate Markov chains, usually with few states (in most cases with two or three states), given the complexity of the estimation procedures, our model is able to involve many "z_t" variables, with multiple states, thanks to the MTD-probit specification as we explain later on.

To be more precise, this paper considers the forecasting of a time series (y_t) that depends on quantitative variable(s) (x_t) and on s discrete or categorical variables, $(S_{1t}, ..., S_{st})$ where S_{jt} $(j = 1, ..., s)$ can take on values in the finite set $\{1, 2, ..., m\}$. We assume that S_{jt} depends on the previous values of $S_{1t-1}, ..., S_{jt-1}, ..., S_{st-1}$, and this dependence is well modeled by a first-order MMC. However, S_{jt} can also depend on some explanatory variables lagged over more than one period - our approach may in fact be viewed as a higher-order MMC (e.g. we may take S_{jt-1} as S_{t-j}, and in this case we would have an s-order Markov chain). We propose using MMC as covariates in a regression model in order to improve the forecast error of a certain dependent variable, provided it is caused, in the Granger sense, by the MMC. Traditionally, and so far, the published literature only addresses the MMC as an end in itself. Here we take advantage of the information about the past state interactions between the MMC categories to forecast the dependent variable more accurately. As far as we know this forecasting problem has not yet been analyzed in the literature.

To form a regression model relating y_t to the categorical variables, we convert the S_{jt} categories into a set of dummy variables as follows:

$$z_{jkt} = 1\left\{S_{jt} = k\right\} \tag{11}$$

Where $1\{.\}$ is the indicator function, $1\left\{S_{jt} = k\right\} = 1$ if $S_{jt} = k$ and 0 otherwise. The proposed methodology also supports the event where S_{jt} is a discrete variable with state space $\{1, 2, ..., m\}$ (say), in which case no dummy variables are needed.

Let us now assume, without any loss of generality, a linear specification like:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\delta} + u_t \tag{12}$$

where:

- $\mathbf{x}_t'$ may be a vector of both deterministic and stochastic components, like AR(1) or other $\mathcal{F}_{t-1}$ or $\mathcal{F}_t$ measurable predetermined terms. Here $\mathcal{F}_t$ represents the information available at time t, i.e. the σ-algebra generated by all events up to time t.
- $\mathbf{z}_t'$ is a vector of dummy variables z_{kjt} , concerning the MMC, defined in (11).
- $\{u_t\}$ is a white noise process mean independent of $\mathbf{x}_t'$ and $\mathbf{z}_t'$.

To forecast y_{t+h} we use the best predictor according to the expected squared forecast error:

$$\mathbf{E}\left[y_{t+h}|\,\mathcal{F}_t\right] = \mathbf{E}\left[\mathbf{x}_{t+h}'\middle|\,\mathcal{F}_t\right]\boldsymbol{\beta} + \mathbf{E}\left[\mathbf{z}_{t+h}'\middle|\,\mathcal{F}_t\right]\boldsymbol{\delta} \tag{13}$$

given the exogeneity of the disturbance term, i.e. $\mathbf{E}\left[u_t|\,\mathcal{F}_{t-1}\right] = 0\ \forall t$.

To illustrate, suppose that we have two categorical variables ($s = 2$) and each categorical datum takes on values in the set $\{1, 2, 3\}$, i.e. $m = 3$. Unwinding the vector $\mathbf{z}_t'$ and the vector $\boldsymbol{\delta}$ it follows that

$$\begin{aligned} y_{t+h} &= \mathbf{x}_{t+h}'\boldsymbol{\beta} + \delta_{11}1\{S_{1t} = 1\} + \delta_{12}1\{S_{1t} = 2\} \\ &+ \delta_{21}1\{S_{2t} = 1\} + \delta_{22}1\{S_{2t} = 2\} + u_t \end{aligned} \tag{14}$$

where S_{jt} represents the $j-th$ categorical series of the MMC (notice that the dummy variable trap is avoided with this specification). Since the values of S_{jt+h} are unknown in the forecasting periods, i.e. for $h \geqslant 1$, we explore possible dependencies between S_{jt+h} and past values of S_{1t+h} and S_{2t+h} using a MMC approach, to predict S_{jt+h}, and consequently, y_{t+h}. If both $S_{1t}$and S_{2t} are discrete variables, the regression equation is simpler:

$$y_{t+h} = x'_{t+h}\beta + \delta_1 S_{1t+h} + \delta_2 S_{2t+h} + u_t. \tag{15}$$

From equations (14) or (15), it is clear that to forecast y_{t+h} one needs to evaluate $P\left(S_{jt+h} = k \middle| \mathcal{F}_t\right)$, for $k = 1, 2, ..., s$. To keep these expressions simple, we make the following assumptions:

Assumption A1. *First order MMC.*

$$P\left(S_{jt} = k \middle| \mathcal{F}_{t-1}\right) = P\left(S_{jt} = k \middle| S_{1t-1} = i_1, \cdots, S_{st-1} = i_s\right). \tag{16}$$

That is, S_{jt} given $\{S_{1t-1}, \cdots, S_{st-1}\}$ is independent of any other variables in $\mathcal{F}_{t-1}$.

Assumption A2. *Homogeneous MMC.*

We have a homogeneous MMC in the sense that

$$P\left(S_{jt} = k \middle| S_{1t-1}, \cdots, S_{st-1}\right) = P\left(S_{jt+h} = k \middle| S_{1t+h-1}, \cdots, S_{st+h-1}\right). \tag{17}$$

Assumption A3. *Contemporaneous needless terms.*

S_{jt} is independent of

$$\left\{S_{1t}, \cdots, S_{j-1t}, S_{j+1t}, \cdots, S_{st}\right\}$$

given $\{S_{1t-1}, \cdots, S_{st-1}\}$, i.e.

$$\begin{aligned} &P\left(S_{jt} = k \middle| S_{1t} = i_1, \cdots, S_{j-1t} = i_{j-1}, S_{j+1t} = i_{j+1}, \cdots,\right. \\ &\left. S_{st} = i_s, S_{1t-1}, \cdots, S_{st-1}\right) = \\ &\qquad = P\left(S_{jt} = k \middle| S_{1t-1}, \cdots, S_{st-1}\right) \end{aligned} \tag{18}$$

To obtain the forecast of y_{t+h} we need to calculate $\mathbf{E}\left[x'_{t+h} \middle| \mathcal{F}_t\right]$ and $\mathbf{E}\left[z'_{t+h} \middle| \mathcal{F}_t\right]$. It is assumed the former expression is known, hence we focus on the latter expression. A generic element of $\mathbf{E}\left[z'_{t+h} \middle| \mathcal{F}_t\right]$ is $\mathbf{E}\left[z_{kj,t+h} \middle| \mathcal{F}_t\right]$ which, by Assumption A1, can be written as

$$\begin{aligned} \mathbf{E}\left[z_{kj,t+h} \middle| \mathcal{F}_t\right] &= P\left(z_{kj,t+h} = 1 \middle| \mathcal{F}_t\right) = P\left(S_{j,t+h} = k \middle| \mathcal{F}_t\right) \\ &= P\left(S_{j,t+h} = k \middle| S_{1t} = i_1, \cdots, S_{st} = i_s\right). \end{aligned} \tag{19}$$

We use the MMC theory to estimate the expression (19), which ultimately leads to the expressions $\mathbf{E}\left[z'_{t+h} \middle| \mathcal{F}_t\right]$ and $\mathbf{E}\left[y_{t+h} | \mathcal{F}_t\right]$. We briefly cover the main aspects of MMC estimation theory in the next section.

3.2 MULTIVARIATE MARKOV CHAINS AS REGRESSORS: MODEL ESTIMATION

In this section we explain our strategy to estimate the parameters defined in equation (12), $\psi=(\beta,\delta)$ and the parameters associated with the multivariate Markov chain, which we denote by η. Let $\theta=(\psi,\eta)$ be the complete vector of parameters, and B and D the parameter space of $\psi=(\beta,\delta)$ and η, respectively. Given the structure of our model and by construction, ψ and η are variation free (see Engle, Hendry, and Richard [31]), since $(\psi,\eta)\in B\times D$, i.e. ψ and η are not subject to cross restrictions so that for any specific admissible value in B for ψ, η can take any value in D. In these circumstances, the conditional distribution of $y_t|\mathbf{S}_t,\mathcal{F}_{t-1}$ depends on ψ only, and the conditional distribution of $\mathbf{S}_t|\mathcal{F}_{t-1}$ depends on η only. In this way the joint density of the complete sample can be sequentially factorized as follows:

$$f\left(y_0,y_1,...,y_n;S_{10},\cdots,S_{1n},\cdots,S_{s0},\cdots,,S_{sn};\theta\right)=$$
$$=\prod_{t=1}^{n}f\left(y_t,\mathbf{S}_t|\mathcal{F}_{t-1};\theta\right)=\prod_{t=1}^{n}f\left(y_t|\mathbf{S}_t,\mathcal{F}_{t-1};\psi\right)\prod_{t=1}^{n}P\left(\mathbf{S}_t|\mathcal{F}_{t-1};\eta\right) \tag{20}$$

Let us focus on $P\left(\mathbf{S}_t|\mathcal{F}_{t-1};\eta\right)=P\left(S_{1t},...,S_{st}|\mathcal{F}_{t-1};\eta\right)$. This expression may be written as:

$$P\left(S_{1t},...,S_{st}|\mathcal{F}_{t-1};\eta\right)=P\left(S_{1t},...,S_{st}|S_{1t-1},...,S_{st-1};\eta\right) \tag{21}$$
$$=\prod_{j=1}^{s}P\left(S_{jt}\middle|S_{1t-1},...,S_{st-1};\eta\right) \tag{22}$$
$$=\prod_{j=1}^{s}P\left(S_{jt}\middle|S_{1t-1},...,S_{st-1};\eta_j\right) \tag{23}$$

where (21) and (22) follow from assumptions A1 and A3, respectively. In equation (23) we decomposed η as $\eta=(\eta_1,...,\eta_s)$, where η_j are the parameters associated with the conditional distribution $S_{jt}\big|S_{1t-1},...,S_{st-1}$. As previously with ψ and η, the vector parameters $\eta_1,...,\eta_s$ are variation free, as will become clear later on. Rearranging all terms one has

$$f\left(y_0,y_1,...,y_n;S_{j0},S_{j1},...,S_{jn};\theta\right)$$
$$=\prod_{t=1}^{n}f\left(y_t|\mathbf{S}_t,\mathcal{F}_{t-1};\psi\right)\prod_{t=1}^{n}\prod_{j=1}^{s}P\left(S_{jt}\middle|S_{1t-1},...,S_{st-1};\eta_j\right)$$
$$=\prod_{t=1}^{n}f\left(y_t|\mathbf{S}_t,\mathcal{F}_{t-1};\psi\right)\prod_{t=1}^{n}P\left(S_{1t}|S_{1t-1},...,S_{st-1};\eta_1\right)$$
$$...\prod_{t=1}^{n}P\left(S_{st}|S_{1t-1},...,S_{st-1};\eta_s\right) \tag{24}$$

and the log likelihood is

$$\log f\left(y_0, y_1, \ldots, y_n; S_{j0}, S_{j1}, \ldots, S_{jn}; \theta\right) = \sum_{t=1}^{n} \log f\left(y_t | \mathbf{S}_t, \mathcal{F}_{t-1}; \psi\right) +$$
$$\sum_{t=1}^{n} \log P\left(S_{1t} | S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_1\right) + \ldots +$$
$$\sum_{t=1}^{n} \log P\left(S_{st} | S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_s\right) \quad (25)$$

This decomposition shows that the parameters can be estimated separately, by maximizing the various expressions in the previous equation, without any loss of consistency or efficiency. Consequently, $\psi = (\beta, \delta)$ is estimated, for example, using the ML in equation (12), and $\boldsymbol{\eta}_j$ $(j = 1, \ldots, s)$ are estimated taking each conditional distribution $S_{jt} \big| S_{1t-1}, \ldots, S_{st-1}$ one at a time, as we will describe in the next section (see for example equation (27)).

3.3 MULTIVARIATE MARKOV CHAIN ESTIMATION

The purpose of this section is to describe a method to estimate the parameters $\boldsymbol{\eta}_j$ defined in the log-likelihood expression (25). As proved in the previous section, the expression

$$\sum_{t=1}^{n} \log P\left(S_{jt} \big| S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_j\right)$$

can be maximized independently of the other terms contained in the log-likelihood function (25). Let

$$P_j\left(i_0 | i_1, \ldots, i_s\right) \equiv P\left(S_{jt} = i_0 \big| S_{1,t-1} = i_1, \ldots, S_{s,t-1} = i_s\right)$$

where $j \in \{1, 2, \ldots, s\}$ and $i_1, \ldots, i_s \in \{1, 2, \ldots, m\}$. It is well known that modelling these probabilities when s and m are relatively large and the sample size is small or even moderate, is unfeasible because the total number of parameters is $m^s(m-1)$, as can be shown. To overcome this problem Raftery [75] considered a simplifying hypothesis for modelling high-order Markov chains (HOMC). Recently, Nicolau [64] proposed an alternative specification, called the MTD-Probit model:

$$\begin{aligned} P_j\left(i_0 | i_1, \ldots, i_s\right) &= P_j^{\Phi}\left(i_0 | i_1, \ldots, i_s\right) \quad (26) \\ &\equiv \frac{\Phi\left(\eta_{j0} + \eta_{j1} P_{j1}\left(i_0 | i_1\right) + \ldots + \eta_{js} P_{js}\left(i_0 | i_s\right)\right)}{\sum_{k=1}^{m} \Phi\left(\eta_{j0} + \eta_{j1} P_{j1}\left(k | i_1\right) + \ldots + \eta_{js} P_{js}\left(k | i_s\right)\right)} \end{aligned}$$

where $\eta_{ji} \in \mathbb{R}$ $(j = 1, ..., s; i = 1, ..., m)$ and Φ is the (cumulative) standard normal distribution function. When S_{jt} is the dependent variable the likelihood is

$$\log L = \sum_{i_1 i_2 \ldots i_s i_0} n_{i_1 i_2 \ldots i_s i_0} \log\left(P_j^{\Phi}\left(i_0 | i_1, ..., i_s\right)\right). \tag{27}$$

and the maximum likelihood estimator is defined, as usual, as $\hat{\eta}_j = \arg\max_{\eta_{j1},...,\eta_{js}} \log L$. The parameters $P_{jk}\left(i_0 | i_1\right)$, $k = 1, , ..., s$ can be estimated in advance, through the consistent estimators $\hat{P}_{jk}\left(i_0 | i_1\right) = \frac{n_{i_1 i_0}}{\sum_{i_0=1}^{n} n_{i_1 i_0}}$ where $n_{i_1 i_0}$ is the number of transitions from $S_{k,t-1} = i_1$ to $S_{jt} = i_0$. This procedure greatly simplifies the estimation procedure and does not alter the consistency of the MLE $\hat{\eta}_j$ estimator, as $\hat{P}_{jk}$ is a consistent estimator of P_{jk}.

3.4 MULTI-STEP FORECAST MODEL

The previous section described how the probabilities

$$P\left(S_{jt} = i_0 \middle| S_{1,t-1} = i_1, ..., S_{s,t-1} = i_s\right)$$

can be estimated. In this section we introduce the h-step-ahead MMC forecast problem, i.e. $P\left(S_{j,t+h} = k \middle| S_{1t} = i_1, \cdots, S_{st} = i_s\right)$. Since we have a homogeneous MMC, the one-step-ahead forecast expression is straightforward, given assumption A2: $P\left(S_{jt+1} = k | S_{1t}, \cdots, S_{st}\right) = P\left(S_{jt} = k | S_{1t-1}, \cdots, S_{st-1}\right)$.

To obtain the h-step-ahead MMC forecast, we consider two procedures. In the first we start to deduce a general formula for the h-step-ahead MMC forecast that can be recursively computed from the previous forecast. Using the discrete version of Chapman-Kolmogorov equations, the formula of total probability, and assumptions A1-A3, we have

$$\begin{aligned}
&P\left(S_{jt+h} = k | S_{1t}, \cdots, S_{st}\right) \\
&= \sum_{i_1}^{m} \sum_{i_2}^{m} \cdots \sum_{i_s}^{m} P\left(S_{jt+h} = k \middle| S_{1t+h-1} = i_1, \cdots, S_{st+h-1} = i_s,\right. \\
&\left. S_{1t}, \cdots, S_{st}\right) \\
&= \sum_{i_1}^{m} \sum_{i_2}^{m} \cdots \sum_{i_s}^{m} P\left(S_{jt+h} = k \middle| S_{1t+h-1} = i_1, \cdots, S_{st+h-1} = i_s\right) \\
&\times \underbrace{P\left(S_{1t+h-1} = i_1 | S_{1t}, \cdots, S_{st}\right)}_{\text{from } h-1} \underbrace{P\left(S_{2t+h-1} = i_2 | S_{1t}, \cdots, S_{st}\right)}_{\text{from } h-1} \\
&\qquad \times \cdots \times \underbrace{P\left(S_{st+h-1} = i_s | S_{1t}, \cdots, S_{st}\right).}_{\text{from } h-1}
\end{aligned} \tag{28}$$

This formula is calculated recursively (notice that it depends on $P\left(S_{jt+h-1}=i_1|S_{1t},\cdots,S_{st}\right), j=1,2,\cdots,s.$). The second procedure is based on the assumption that

$$P\left(S_{jt+h}=i_0\middle| S_{1,t}=i_1,...,S_{s,t}=i_s\right)= \frac{\Phi\left[\eta_{j0}+\eta_{j1}P\left(S_{jt+h}=i_0\middle| S_{1t}=i_1\right)+ +...+\eta_{js}P\left(S_{jt+h}=i_0\middle| S_{st}=i_s\right)\right]}{\sum_{k=1}^{m}\Phi\left[\eta_{j0}+\eta_{j1}P\left(S_{jt+h}=k\middle| S_{1t}=i_1\right)+ +...+\eta_{js}P\left(S_{jt+h}=k\middle| S_{st}=i_s\right)\right]} \quad (29)$$

which is clearly a natural extension of equation (40). This expression requires that $P\left(S_{jt+h}=i_0\middle| S_{kt}=i_k\right)$ be computed in advance. From the Chapman-Kolmogorov equations and the formula of total probability, it can be easily seen that

$$P\left(S_{jt+h}=i_0\middle| S_{kt}=i_k\right)= \sum_{\alpha=1}^{m}P\left(S_{jt+h}=i_0\middle| S_{k,t+h-1}=\alpha\right)P\left(S_{k,t+h-1}=\alpha| S_{kt}=i_k\right). \quad (30)$$

This expression is equal to the element (i_0,i_k) of the matrix product $P^{(jk)}\left(P^{(kk)}\right)^{h-1}$ where $P^{(jk)}$is a matrix with elements

$$P\left(S_{jt}=i_0\middle| S_{kt-1}=i_k\right).$$

We found formula (29) computationally easier to implement than (28).

We may now establish the algorithm behind the forecast of y_{t+h} :

1. Run the regression model $y_t = x_t'\beta+z_t'\delta+u_t$ and estimate β and δ using the OLS method or any other method.
2. Obtain $\hat{\eta}_j=\arg\max_{\eta_{j1},\dots,\eta_{js}}\log L$ where the log-likelihood refers to equation (27).
3. From the estimates $\hat{\eta}_j$ calculate $P\left(S_{jt+1}=k|S_{1t},\cdots,S_{st}\right)$, and derive the expressions $P\left(S_{jt+h}=k|S_{1t},\cdots,S_{st}\right)$, either recursively from formula (28) or from formula (29).
4. Finally, obtain the forecast y_{t+h} by calculating

$$\mathbf{E}\left[y_{t+h}|\mathcal{F}_t\right]=\mathbf{E}\left[x_{t+h}'\middle|\mathcal{F}_t\right]\beta+\mathbf{E}\left[z_{t+h}'\middle|\mathcal{F}_t\right]\delta \quad (31)$$

3.5 MONTE CARLO SIMULATION STUDY

3.5.1 *Monte Carlo Simulation Study: Procedure and Design*

In this section we evaluate the MMC predictive potential through a Monte Carlo simulation problem. The goal is to construct a model where the MMC, transformed into $s\times(m-1)$ dummy variables (one dummy for each state minus one, for each category), play the role of covariates, seeking to gauge how they help forecast a certain dependent variable.

We consider here a simple process with two categories ($s = 2$) with each one taking values of 1, 2 or 3 ($m = 3$). We simulate the MMC in accordance with the following algorithm:

1. Initialize the process $\{(S_{1t}, S_{2t})\}$ by assigning arbitrary values for S_{10} and for S_{20}.
2. Define two $m^s \times m$ TPMs whose elements are, respectively, the following probabilities

$$\begin{aligned} &P\left(S_{1t} = i_o | S_{1t-1} = i_1, S_{2t-1} = i_2\right) \\ &P\left(S_{2t} = i_o | S_{1t-1} = i_1, S_{2t-1} = i_2\right) \end{aligned} \tag{32}$$

   (see the definition of the data-generation process below)
3. Given the initial values S_{10} and S_{20} (step 1), simulate the multivariate process $\{(S_{1t}, S_{2t})\}$, $t = 1, .., T$ as follows:
   a) simulate U_1, uniformly distributed on [0, 1];
   b) let us define $p_i^{[1]} \equiv P\left(S_{1t} = i | S_{1t-1} = i_1, S_{2t-1} = i_2\right)$;
   c) assign a value to S_{1t} according to the rule:

$$S_{1t} = \begin{cases} 1 & \text{if} \quad 0 \leqslant U_1 < p_1^{[1]} \\ 2 & \text{if} \quad p_1 \leqslant U_1 < p_1^{[1]} + p_2^{[1]} \\ 3 & \text{if} \quad p_1 + p_2 \leqslant U_1 < 1 \end{cases}$$

   d) repeat this procedure for S_{2t} (using $U_2 \sim U(0, 1)$, independent of U_1) .
4. Repeat the steps 1-4 until $t = T$.

Thus, we construct our 4 dummy variables, as in (11), such that: $z_{jk,t} = 1\{S_{jt} = k\}$, $k = 1, \cdots, m-1$.

We consider the following linear data-generation process (DGP) where

- $z_t' \equiv \left(\begin{array}{cccc} z_{11} & z_{12} & z_{21} & z_{22} \end{array}\right)$, $\delta = \left(\begin{array}{cccc} 1 & 1 & 1 & 1 \end{array}\right)'$, for simplicity,
- $x_t' = \left(\begin{array}{cc} 1 & x_t \end{array}\right)$ and x_t (such as u_t) is i.i.d. $N(0, 1)$, $\beta = \left(\begin{array}{cc} 1 & 1 \end{array}\right)'$.

To fully define the DGP, we arbitrarily construct the TPM as follows:

For convenience, we assume that S_{1t} and S_{2t} have the same transition probabilities, i. e. that $P\left(S_{2t} = i_0 | S_{1t-1} = i_1, S_{2t-1} = i_2\right) = P\left(S_{1t} = i_0 | S_{1t-1} = i_1, S_{2t-1} = i_2\right)$.

We aim to compare the dependent variable h-step-ahead forecast errors produced by four different hypotheses:

- **Case 1**: The values of dummy variables at $t + h$ are known,

$$\hat{z}_{jkt+h}^{(1)} = z_{jkt+h} \tag{33}$$

S_{1t-1}		1	1	1	2	2	2	3	3	3
S_{2t-1}		1	2	3	1	2	3	1	2	3
S_{1t}	1	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1
	2	0.1	0.4	0.1	0.4	0.8	0.1	0.1	0.4	0.1
	3	0.8	0.1	0.8	0.1	0.1	0.4	0.8	0.1	0.8

Table 5: Transition Probability Matrix

- **Case 2**: The values of dummy variables at $t+h$ are predicted using the following proposed methodology

$$\hat{z}^{(2)}_{jkt+h} = \hat{P}\left(S_{jt+h} = k \middle| S_{1t} = i_1, S_{2t} = i_2\right) \tag{34}$$

  where $\hat{P}\left(S_{jt+h} = k \middle| S_{1t} = i_1, S_{2t} = i_2\right)$ is obtained according to expression (30).

- **Case 3**: The values of dummy variables at $t+h$ are predicted using marginal means

$$\hat{z}^{(3)}_{jkt+h} = T^{-1} \sum_{t=1}^{T} z_{jkt}. \tag{35}$$

  Note that we estimate the event $\mathcal{I}_{\{S_{jt}=k\}}$ using a consistent estimator for the marginal probability, $P\left(S_{jt+h} = k\right)$.

- **Case 4**: The dummies are omitted, i.e $\hat{z}^{(4)}_{jkt+h} \equiv 0$.

Out-of-sample forecasts were generated by the so-called recursive (expanding windows) forecasts. An initial sample using data from $t = 1$ to $T = 1000$ is used to estimate the models, and h-step ahead out-of-sample forecasts are produced starting at time $T = 1000$. The sample is increased by one, the models are re-estimated, and h-step ahead forecasts are produced starting at $T+1$. This procedure is repeated 1000 times, i.e. we considered 1000 out-of-sample forecasts, and the forecasting time horizon was defined as $h = 1, 2, 3, 4, 5$. Lastly, we assessed the quality of the forecast using the statistics $MSE_{lh} = N^{-1} \sum_{t=T}^{T+999} \hat{e}^2_{l,t+h}$ where $N = 1000$ is the number of replicas considered in the experiment and $e_{l,t+h}$ is the forecast error produced by model l $(l = 1, 2, 3, 4)$ at the $h-th$ forecast step, i.e.

$$e_{l,t+h} \equiv y_{t+h} - \hat{y}^{(l)}_{t+h} \tag{36}$$

where

$$\hat{y}^{(l)}_{t+h} \equiv \mathbf{x}'_{t+h}\hat{\boldsymbol{\beta}} + \hat{\mathbf{z}}^{(l)'}_{t+h}\hat{\boldsymbol{\delta}}$$

and $\hat{\mathbf{z}}^{(l)'}_{t+h} \equiv \begin{pmatrix} z^{(l)}_{11t+h} & z^{(l)}_{12t+h} & z^{(l)}_{21t+h} & z^{(l)}_{22t+h} \end{pmatrix}$, for $l = 1, 2, 3, 4$.

3.5.2 *Monte Carlo Simulation Study: Discussion of Results*

In this section we report the results of the Monte Carlo study presented in the previous section, investigating the potential forecast gains of a dependent variable, derived by processing categorical interrelated regressors as a MMC, i.e. by exploiting intra and inter-transition probabilities between categorical regressors. Figure 1 presents the MSE_{lh} for $l = 1, 2, 3, 4$ and $h = 1, 2, 3, 4$.

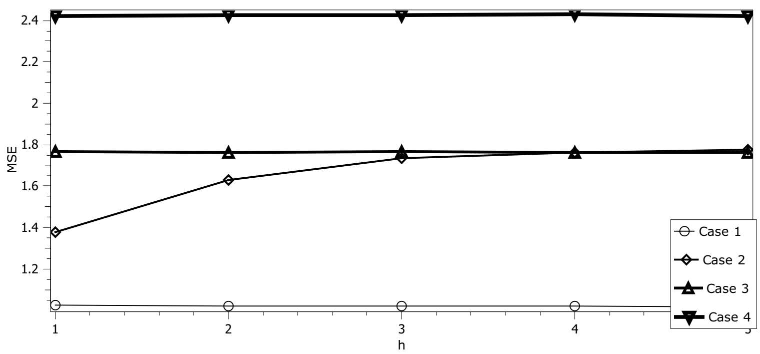


Figure 1: Results of the forecast errors MSE_{lh}

As expected, case 1 presents the best results, since the forecast of y_{t+h} is based on the actual values z_{jkt+h}, and case 4 the worst result, since the dummies z_{jkt+h} were simply ignored. Case 2 uses the proposed methodology, and hence explores the intra and inter-transition probabilities between categorical regressors; it clearly produces better results than case 3, in which the forecasts are based on the estimate of the marginal probabilities $P\left(S_{jt+h} = k\right)$. To confirm the advantage of the proposed method over the marginal probabilities we carried out the Diebold and Mariano [28] (DM) test, that allows us to assess the significance of the MSE difference between those models. As is known, the DM can be trivially calculated by regression $\hat{e}^2_{3,t+h} - \hat{e}^2_{2,t+h}$ on an intercept, using heteroskedasticity and autocorrelation robust (HAC) standard errors. Our results (available upon request) show that the proposed model outperforms the forecasts based on the marginal mean for $h = 1$ and $h = 2$ (p-value zero), and possibly $h = 3$ (p-value 0.08). When h increases, the advantage of using our model dissipates, which is to be expected taking into account that in stationarity and weak dependence assumptions, the conditional probabilities converge into the stationary probabilities, i.e. $P\left(S_{jt+h} = i_0 \middle| S_{1t} = i_1, S_{2t} = i_2\right) \to P\left(S_{jt} = i_0\right)$ as $h \to \infty$.

3.6 conclusions

This paper proposed a new concept by using MMC as covariates in a regression model in order to improve the forecast error of a certain dependent variable, provided it is caused by the MMC. Traditionally, the published literature only addresses the MMC as an end in itself. In a context of an endogenous variable that depends on some time-dependent categorical or discrete variables, we show that taking advantage of the information about the past state interactions between the categorical variables through a MMC specification via modelling the intra and inter-transition probabilities within and between data categories, may lead to a substantial forecasting improvement of that endogenous variable, as the Monte Carlo experiment has shown.

4

THE CHANGING ECONOMIC REGIMES AND EXPECTED TIME TO RECOVER OF THE PERIPHERAL COUNTRIES UNDER THE EURO: A NONPARAMETRIC APPROACH

Note: This article has been published in co-authorship with Francisco Louçã and João Nicolau in Physica A: Statistical Mechanics and its Applications

4.1 INTRODUCTION

The expected time to cross a given threshold is an important concept in stochastic analysis, although not commonly used in economic investigations. In the case of this paper, we develop a new method [65] to compute the expected time to recover or to adapt from a negative or positive shock or change. Independently of considerations on the endogenous or exogenous nature of perturbations in the dynamics of the aggregate measure of economic activity, the GDP, and accepting for the purpose of the computation the approximation provided by the required two assumtpions (Markovian property and transformation for stationarization of data), we apply this method to appreciate the impact of the monetary regime change occurring from 1999 on the dynamics of the economies of the peripheral countries in Europe. Section 2 summarizes the methodology and Section 3 the empirical results, whereas Section 4 presents a conclusion.

4.2 METHODOLOGY: A NONPARAMETRIC METHOD TO ESTIMATE THE EXPECTED TIME

The expected time for the economy to recover after a slump is an important indicator on how robust the economy is to shocks and how effective the policies and institutions are to regain the path of *normal growth*.

The econometric literature offers few alternative approaches to analyse this issue. A possible measure, commonly related to the level of persistence of a time series, is the half-life which is usually defined as the number of periods required for the impulse response to a unit shock of a time series to dissipate by half. However, empirical studies of half-lives have documented some issues related to the precision and unbiasedness of the estimates [61]. Most of the problems are related to incorrect model specification (apart from other sources such as temporal aggregation, structural breaks, etc.). Furthermore, half-life implies that a positive and negative shock of equal magnitude has the same impact on the impulse response function; however, the reversion to a fixed point (e.g. stationary mean) may display different behaviour depending on whether the process is below or above that point.

Another way to discuss the time to recover can be based on the concept of expected time (ET) to cross some thresholds. For example, suppose that the GDP growth rate crosses some negative value say x_0, indicating that the economy is in recession; then define a higher level or threshold that the process eventually reaches in the future, say x_1. The expected time for the process to go from x_0 to x_1 is an indication of how resilient and robust the economy is to recover from recession. The ET concept has received little attention in economics. One of the reasons is probably the difficulty in obtaining a simple procedure to calculate, for example, the ET to reach a threshold. In fact, analytical results on first hitting time problems (from which expected time may be calculated) are mostly based on stochastic processes of diffusion type or Markov chains where explicit analytical expressions are usually available. First hitting times are often used in mathematical finance, biology and other life sciences, where Markov chains and stochastic differential equations are more commonly used, for example, to study time to extinction or default (in finance). Nonetheless, ET may also be a very useful tool in economics to discuss topics such as the speed of mean-reversion, the time to equilibrium, and especially in the current case the time to recovery.

In this paper, we use a new estimator by Nicolau [65] to estimate the expected time to cross some thresholds. This estimator is formulated in a completely nonparametric framework and uses only two assumptions: Markovian property and stationarity. Standard errors can also be computed. We sketched the main ideas of the method here.

Let y be the GDP growth process with state space $\mathbb{R}$. We assume that:

Assumption A4. y *is a Markov process of order* r *(*$1 \leqslant r < \infty$*),*

and

Assumption A5. y *is a strictly stationary process.*

Under assumption A5, it can be proved that starting the process from a level a not belonging to the generic set A, the process y visits A an infinite number of times as $t \to \infty$, almost surely, see Meyn and Tweedie [57, chap. 9]. This property is of course crucial for (pointwise) identification.

We consider the hitting time $T \equiv T_{x_1} = \min\{t > 0 : y_t \geqslant x_1\}$ and suppose that the process starts at value $x_0 < x_1$. The case $x_0 > x_1$ with $T_{x_1} = \min\{t > 0 : y_t \leqslant x_1\}$ is almost analogous. A brief remark on this case will be made later on. The distribution of T is usually difficult to deduce for general non-linear processes. However, there is a simple nonparametric method to estimate these quantities. Set

$S_0 = 1$ if $y_0 = x_0$ (note that the process starts at $y_0 = x_0$). Now define the following transformation for $k \geqslant 0$

$$S_t = \begin{cases} 1 & \text{if } y_t < x_1, y_{t-1} < x_1, ..., y_{t-k+1} < x_1, y_{t-k} \leqslant x_0 \\ 2 & \text{if } x_0 < y_t \leqslant x_1, x_0 < y_{t-1} \leqslant x_1, ..., x_0 < y_{t-k+1} \leqslant x_1, y_{t-k} \geqslant x_1 \\ 3 & \text{otherwise.} \end{cases} \tag{37}$$

Figure 2 illustrates the map (37) for a hypothetical trajectory of y.

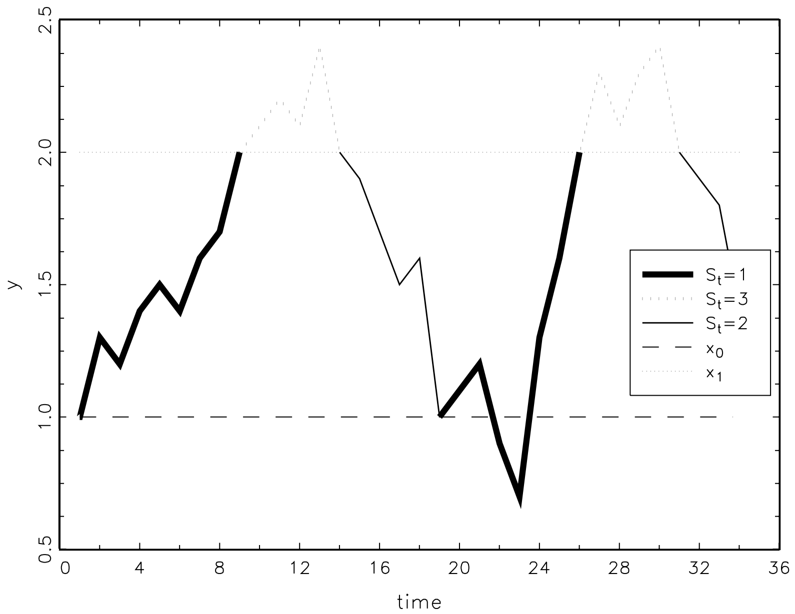


Figure 2: Illustrating map (37), where $x_0 = 1$, $x_1 = 2$. Thick line: $S_t = 1$; thin line: $S_t = 2$; dot line: $S_t = 3$

The probabilities of T, which can be difficult or impossible to obtain from y, may be easily calculated from process S_t. It can be proved that

$$P(T = t) = (1 - p_t) \prod_{i=1}^{t-1} p_i = (1 - p_t) p_{t-1} p_{t-2} ... p_1 \tag{38}$$

where $p_t = P(S_t = 1 | S_{t-1} = 1, S_{t-2} = 1, ..., S_0 = 1)$. Our strategy is to treat S_t as a Markov chain with state space $\{1, 2, 3\}$ from which we then estimate the relevant parameters. The following result supports our approach.

Propositon 1. *Suppose that y is a rth order Markov process. Then S is a rth order Markov chain.*

From the A4 assumption and previous proposition, one has $p_t = P(S_t = 1 | S_{t-1} = 1, S_{t-2} = 1, ..., S_{t-r} = 1)$. The probabilities p_t can be estimated from standard Markov chain inference theory.

We first analyse the case $r = 1$. To emphasise the dependence of S_t on the thresholds x_0, x_1, we write the transition probability matrix as $\mathbf{P}(x_0, x_1) = \left[P_{ij}(x_0, x_1)\right]_{3\times 3}$ where $P_{ij} = P_{ij}(x_0, x_1) \equiv P(S_t = j | S_{t-1} = i)$. The only parameter of interest is P_{11}. If S is a first order Markov chain, i.e. $r = 1$, then $p_t = P(S_t = 1 | S_{t-1} = 1) = P_{11}$ and

$$\mathbf{E}[T] = \sum_{t=1}^{\infty} t p_t = (1 - P_{11}) \sum_{t=1}^{\infty} t P_{11}^{t-1} = \frac{1}{1 - P_{11}}. \tag{39}$$

This quantity can be easily estimated from the maximum likelihood estimate $\hat{P}_{11} = n_{11}/n_1$ where n_{11} is the number of transitions of type $S_{t-1} = 1$, $S_t = 1$ and n_1 counts the number of ones (i.e. $S_t = 1$) in the sample.

Propositon 2. *We have* $\hat{P}_{11} = n_{11}/n_1 \xrightarrow{p} P_{11}$ *and* $\sqrt{n}\left(\hat{P}_{11} - P_{11}\right) \xrightarrow{d} N(0, P_{11}(1 - P_{11})/\pi_1)$ *where* π_1 *is such that* $n_1/n \xrightarrow{p} \pi_1$.

It is interesting to observe that the process y has to visit (or cross) the threshold x_1 an infinite number of times over time, in order to achieve consistency, and this follows from A5, and in particular from positive Harris recurrence of y and also of S. This prevents, for example, having only ones in the sequence of S (which represents the case where y never visits x_1) and consequently $\mathbf{E}[T] = \infty$.

Propositon 3. *Let* $\widehat{\mathbf{E}[T]} = 1/\left(1 - \hat{P}_{11}\right)$. *We have in the case* $r = 1$

$$\begin{aligned}
&\widehat{\mathbf{E}[T]} \xrightarrow{p} \mathbf{E}[T], \\
&\sqrt{n}\left(\widehat{\mathbf{E}[T]} - \mathbf{E}[T]\right) \xrightarrow{d} N\left(0, \frac{P_{11}}{(1 - P_{11})^3 \pi_1}\right), \\
&0 < P_{11} < 1.
\end{aligned}$$

Propositon 4. *Let us now focus on the case* $r > 1$. *We saw previously that*

$$P(T = t) = (1 - p_t) \prod_{i=1}^{t-1} p_i$$

where $p_t = P(S_t = 1 | S_{t-1} = 1, S_{t-2} = 1, ..., S_0 = 1)$. *Given that* $p_t = p_r$ *if* $t > r$, *in view of the Markovian property, we have*

$$P(T = t) = \begin{cases} (1 - p_t) \prod_{i=1}^{t-1} p_i & t \leqslant r \\ \left((1 - p_r) \prod_{i=1}^{r-1} p_i\right) p_r^{t-r} & t > r. \end{cases}$$

Consequently, we have

$$\begin{aligned}
\mathbf{E}[T] &= \sum_{t=1}^{r} t(1-p_t)\prod_{i=1}^{t-1} p_i + \left((1-p_r)\prod_{i=1}^{r-1} p_i\right)\sum_{t=r+1}^{\infty} t p_r^{t-r} \\
&= \sum_{t=1}^{r} t(1-p_t)\prod_{i=1}^{t-1} p_i + \left((1-p_r)\prod_{i=1}^{r-1} p_i\right)\times \\
&\times \frac{p_r(1+r-rp_r)}{(1-p_r)^2}. \qquad (40)
\end{aligned}$$

This expression simplifies to the following formulas:

$$\begin{aligned}
r &= 1 \Rightarrow \mathbf{E}[T] = \frac{1}{1-p_1} = \frac{1}{1-P_{11}} \text{ (see equation 39)}, \\
r &= 2 \Rightarrow \mathbf{E}[T] = \frac{1+p_1-p_2}{1-p_2}, \text{ etc.}
\end{aligned}$$

Since the Markov chain is homogeneous, it follows that

$$\begin{aligned}
p_k &= P(S_k = 1 | S_{k-1} = 1, ..., S_0 = 1) \\
&= P(S_t = 1 | S_{t-1} = 1, ..., S_{t-k} = 1),
\end{aligned}$$

$k < r$, and in particular, $p_1 \equiv P(S_1 = 1 | S_0 = 1) = P_{11}$. To estimate p_k we use the maximum likelihood estimate $\hat{p}_k = A/B$, where A is the number of transitions from $S_{t-1} = 1, ..., S_{t-k} = 1$ to $S_t = 1$ and B is the number of cases where $S_{t-1} = 1, ..., S_{t-k} = 1$. However, modelling these probabilities may be problematic when k is relatively large and the sample size is small. Therefore k should be no than 4 or 5 (say), depending on the sample size, the level of persistence of y and the thresholds x_0 and x_1. Nevertheless, the literature provides methods to deal with higher k, for example by using the Mixture Transition Distribution [75] or the probit-Mixture Transmodellingition Distribution [26, 65].

We must also make a few observations on the statistical inference in the case $r > 1$. The estimate of $\mathbf{E}[T]$ is straightforward: one needs only to replace the unknown parameters with the corresponding ML estimates. The estimator thus obtained is obviously consistent for $\mathbf{E}[T]$. However, as it is evident from (40), an exact asymptotic expression for the distribution of $\widehat{\mathbf{E}[T]}$ is difficult to obtain. To overcome this issue, we consider the regeneration-based bootstrap procedure of Athreya and Fuh [5] (see also Nicolau [65] for more details).

4.3 EMPIRICAL ILLUSTRATION: THE CHANGING ECONOMIC REGIMES OF THE PERIPHERAL COUNTRIES UNDER THE EURO

4.3.1 *The economic problem*

The investigation on economic fluctuations dominated macroeconomics through the first half of the twentieth century but was thereafter declared obsolete by an over-reaching confidence in stabilization policies. In this sense, Paul Samuelson joked at the fiftieth anniversary conference of the U.S. National Bureau of Economic Research, a major center for business cycle research, that its success was putting the organization out of a job. Nevertheless, the major recessions of the end of the century and that ignited by the subprime crash demonstrated major fragilities both in the economic structure of the developed countries and in their economic prescriptions and models.

The revival of business cycle analysis is built on different theoretical contributions, from the traditional approaches by Mitchell [59] and Schumpeter [80] to the literature on long term processes of match and mismatch between the techno-economic paradigm and the institutional framework [33], to the historic analysis of different epochs [3, 97], to the interpretation of the articulation among different institutional and economic factors [16, 17, 49], the convergence of general purpose technologies [50] and, finally, to the discussion on secular stagnation [16, 29, 46, 90].

Although considering these contributions, we concentrate in this paper in an empirically oriented investigation in order to detect major structural changes in the schedule of quarterly GDP for some European countries (namely on Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and UK) through the long period of 1962-2016, suggesting that they indicate a regime change for the economies under scrutiny after the creation of the Monetary Union. In this case, therefore, the analysis of business cycles and long term dynamics is combined with the interpretation of the effect of a crucial change in the monetary regime of these economies.

Instead, the prediction suggesting that the Euro would imply the real convergence of the different economies has been canonical among the proponents of the single currency. Following the theory of an optimal currency zone, the deregulation procedures and the free movement of goods, capital and labor would allow for leveling the interest and the implicit exchange rates and to the convergence of factor prices in the different economies. As expressed by the then governor of the Bank of Portugal and current vice-president of the ECB, Vitor Constâncio, in a sworning-in 2000 speech, *without a currency of our own, we shall never again face the same balance of payments problems of the past. There is no macroeconomic monetary problem and no restrictive measures need to be taken for balance of payments reasons. No one analyses the macro size of the external account of the Mississippi or of any other region belonging to a large monetary union.*[1] Recently, many authors challenged this view, considering the experience [2, 46, 86, 87].

1 Available at:https://www.bportugal.pt/en-US/OBancoeoEurosistema/IntervencoesPublicas/Pages/intervpub20000223.aspx

	1962-1998	1999-2016
$\sigma_{(\bar{y}_{ik}-\hat{\sigma}_{y_{ik}})}$	2.27	4.00
$\sigma_{(\bar{y}_{ik}+\hat{\sigma}_{y_{ik}})}$	1.70	4.23
$\sigma^{EM}_{(\bar{y}_{ik}-\sigma_{y_{ik}})}$	2.49	4.68
$\sigma^{EM}_{(\bar{y}_{ik}+\sigma_{y_{ik}})}$	1.96	4.96

Table 6: Sample Standard Deviation of Expected Reversion Time

4.3.2 *Empirical estimation*

From a practical perspective, we estimate ET for the different starting points x_0, but the same threshold x_1. We call these estimates the ET curve or ETC. The value x_1 is defined as $x_1 = \bar{y}$ (empirical mean of y) which is the best estimate of the stationary mean. Therefore, the ETC measures the expected time for the process to revert to its stationary mean.

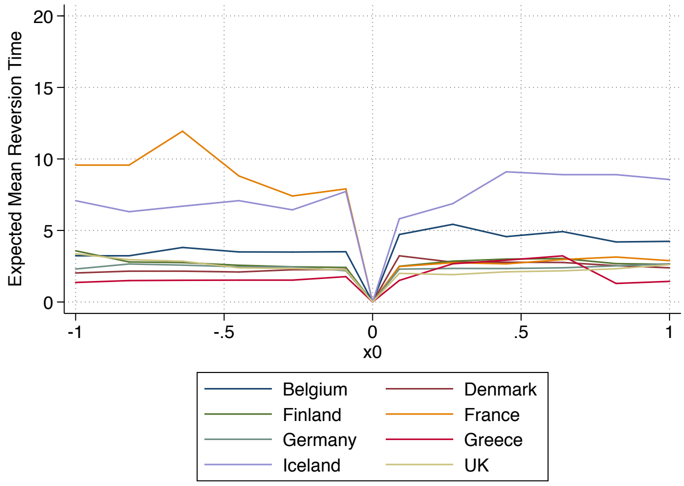
It should be point out the fact that, given the nonstationary nature of GDP process, we considered log-differenciated series in order to achieve stationarity (Assumption A5) as it is confirmed by the aumented Dickey-Fuller tests.

To obtain the ETC we generate different values for x_0 equally spaced between $\bar{y} - \delta\hat{\sigma}_y$ and $\bar{y} + \delta\hat{\sigma}_y$, where $\hat{\sigma}_y$ is the sample standard deviation of y and δ is a parameter that controls the amplitude of the interval $(\bar{y} - \delta\hat{\sigma}_y, \bar{y} + \delta\hat{\sigma}_y)$. In our analysis we set $\delta = 1$. In order to compare the different ETC we standardize our data, so that all the standardized GDP will have zero mean and unit variance (hence, the point x_1 turns out to be zero). We have considered a second order Markov process (i.e. $r = 2$) based on the partial autocorrelation of y (we have analyzed seasonal adjusted series). However, it should be mentioned that other values of r lead to approximately the same results.

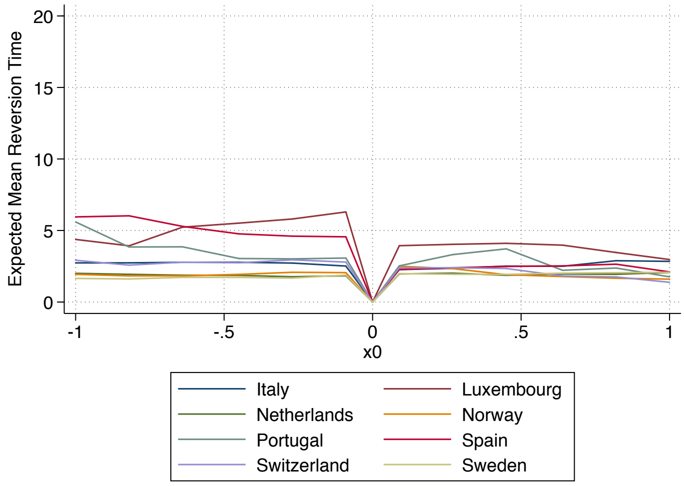
Figures 3a and 3b display the Expected Time Curves for the 1962-1998 period and Figures 4a and 4b show the ETC for the 1999-2016 period on the GDP growth rates for 16 European countries. It is interesting to note that the dispersion of the mean expected reversion time among countries is much higher in 1999-2016 than in the previous period (1962-1998), both for negative and positive deviations, as we shall see next.

Figure 5 displays the scatter plot of the average growth rates against the expected mean reversion time given a positive deviation of

$$(\bar{y}_{ik} + \sigma_{y_{ik}}),$$



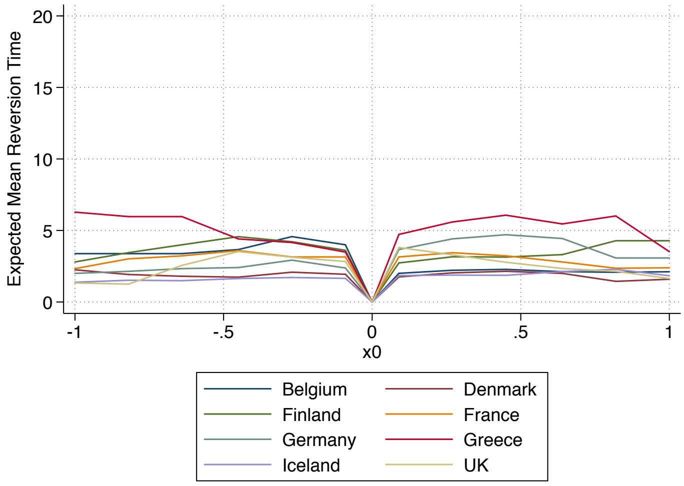
(a)



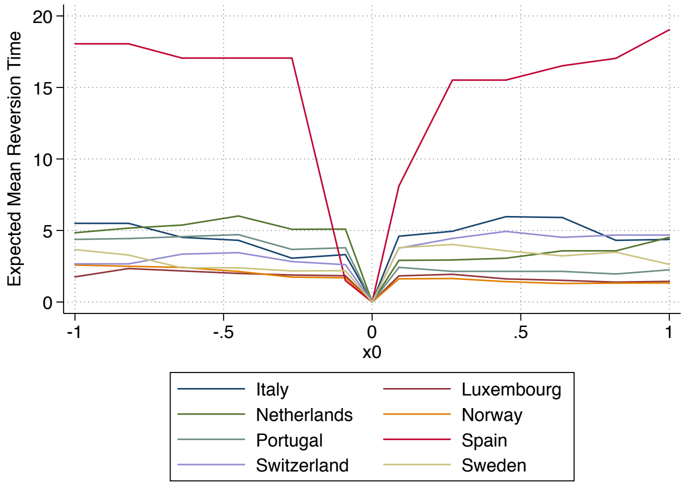
(b)

Figure 3: ETC for the 1962-1998 period

where $\bar{y}_{ik}$ is the average growth rate of the $i-th$ country at period k ($i = 1, 2, ..., 16$ and $k = 1, 2$), for each subsample, and Figure 6 is similar to Figure 5 but represents the expected mean reversion time given a negative deviation of $(\bar{y}_{ik} - \sigma_{y_{ik}})$. Reference red lines denote the medians (considering the entire sample), therefore the Figure is divided into 4 quadrants. Regarding Figure 6, while the second quadrant (top left) represents the best possible situation (high growth rates and small mean reversion time given negative deviations) the fourth quadrant (bottom right) depicts the worst case scenario (small growth rates and high mean reversion time given negative deviations). On the one hand, after 1999, all the southern countries, notably, Portugal, Italy, Greece and Spain have moved form first or second quadrant to the fourth one. This circumstance is confirmed by the dendograms



(a)



(b)

Figure 4: ETC for the 1999-2016 period

(Figures 8 and 12) where the southern European countries form a well defined cluster after 1999 onwards, whilst before 1999 (Figures 7 and 11) the southern countries integrate different clusters among themselves. Nevertheless, one can observe a generalized mass migration of countries to the third and fourth quadrants of the scatter plots motivated by higher ET and lower growth rates.

However, the generality of the non Euro member states occupied the first quadrant until 1999 and moved to the third one quadrant after 1999 (there are no non Euro member states in the fourth quadrant) which means low recovery times after negative deviations in comparison with the Euro states. Moreover, the ET of the non Euro economies had not change a great deal from the 1962-1998 period to the 1999-2016.

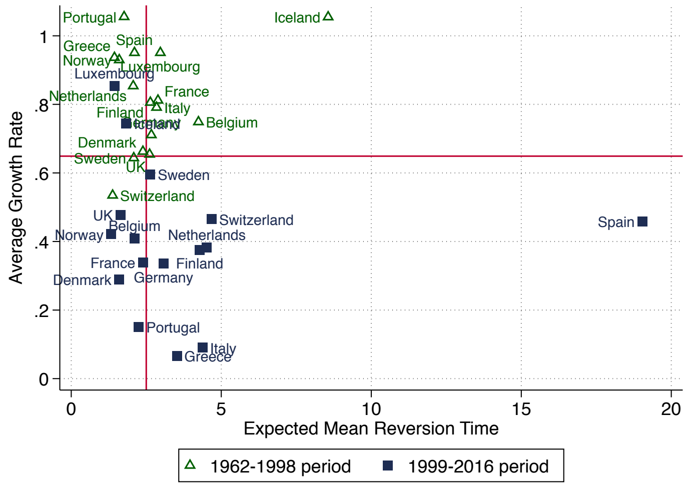


Figure 5: Scatter plot for positive deviations from the mean

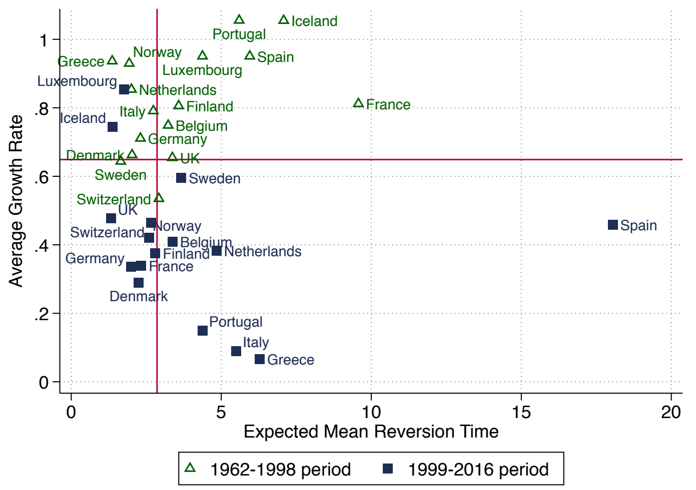


Figure 6: Scatter plot for negative deviations from the mean

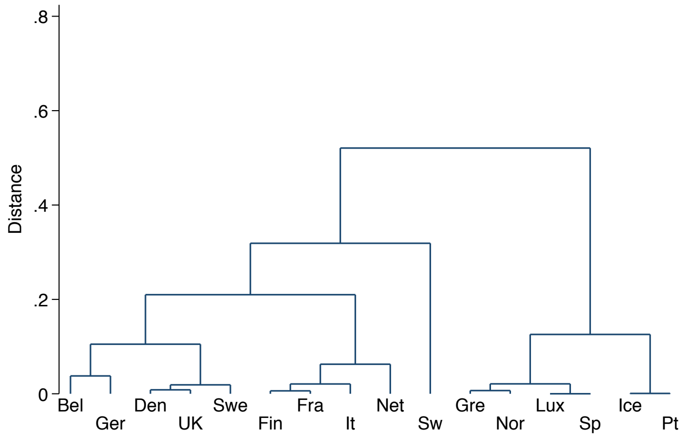


Figure 7: Dendogram: average growth rates for the 1962-1998 period

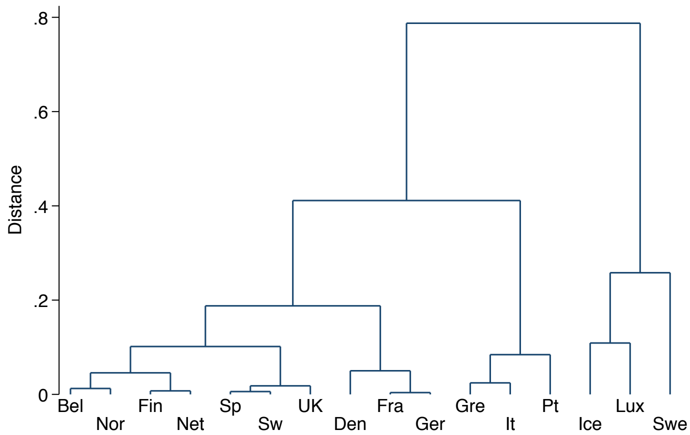


Figure 8: Dendogram: average growth rates for the 1999-2016 period

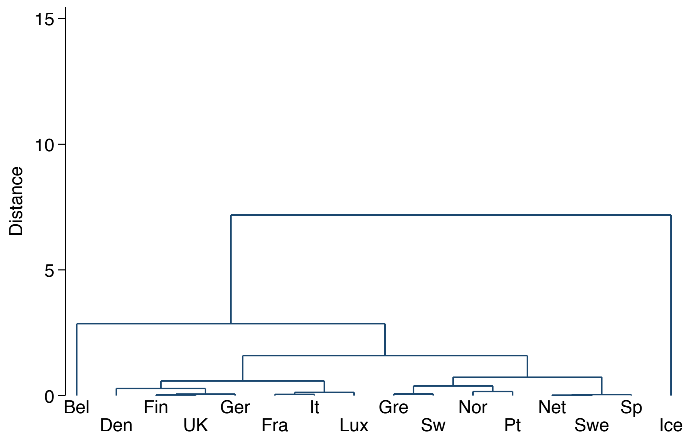


Figure 9: Dendogram: positive deviations from the mean 1962-1998 period

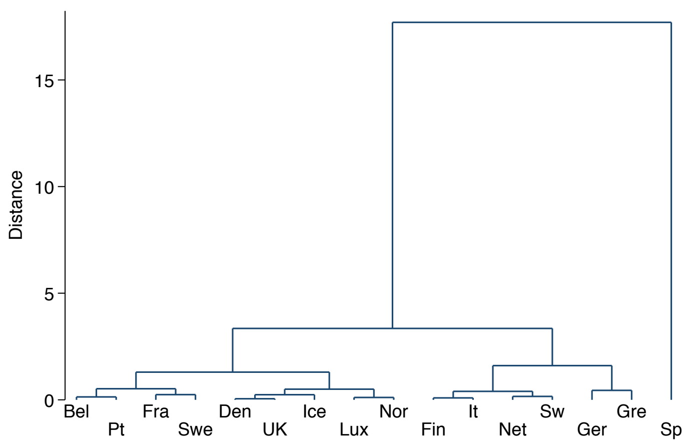


Figure 10: Dendogram: positive deviations from the mean 1999-2016 period

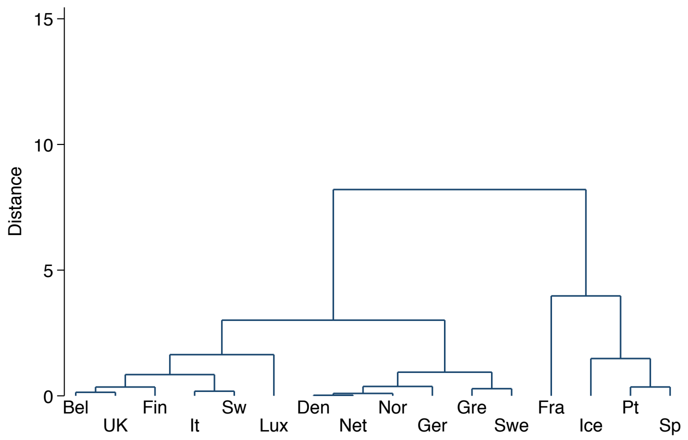


Figure 11: Dendogram: negative deviations from the mean 1962-1998 period

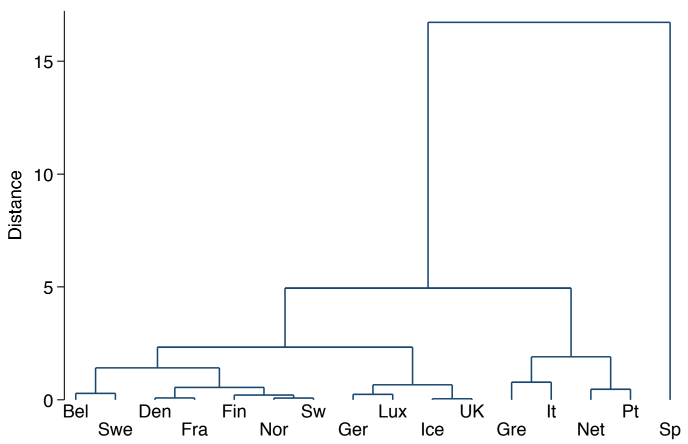


Figure 12: Dendogram: negative deviations from the mean 1999-2016 period

On the other hand, after 1999 the behavior of the countries is much asymmetrical, both in terms of growth rates and in terms of (positive and negative) mean reversion time, with respect to the Euro member countries, and in particular regarding the Southern economies. In fact, unlike the 1962-1998 period, from 1999 countries exhibits a high degree of dispersion since the points are substantially diffuse over the clouds (Figures 5 and 6 and Table 6).

The two extreme points of the ETC represent the expected reversion time when $x_0 = -1$ and $x_0 = 1$, that is to say, the mean ET to recovery given a deviation of $\bar{y}_{ik} - \hat{\sigma}_{y_{ik}}$. Table 6 displays the standard deviation of the ET for positive deviations ($x_0 = 1$) and for negative deviations ($x_0 = -1$) between countries, for the whole set of countries - $\sigma_{(\bar{y}_{ik} - \hat{\sigma}_{y_{ik}})}$. In fact, the standard deviation of expected reversion time given negative deviations raised more than 76% from 2.27 to 4.00 while the positive counterpart more than doubled (raised form 1.70 to 4.23) highlighting the overdispersion behaviour of the recovery ET after 1999.

In order to analyse the *Euro effect*, let $\sigma^{EM}_{(\bar{y}_{ik} - \hat{\sigma}_{y_{ik}})}$ and $\sigma^{EM}_{(\bar{y}_{ik} + \hat{\sigma}_{y_{ik}})}$ denote, respectively, the mean reversion time given negative and positive deviations form the mean of one standard deviation for the Euro member countries only (excluding Denmark, Norway, Switzerland, UK and Sweden). This asymmetrical overdispersion pattern of the mean expected reversion time among countries is also much higher from 1999-2016 than in 1962-1998 for the Euro member countries comparing with the whole set of countries. Considering only Euro member states, the relative dispersion (1999-2016 vs 1962-1998) of the expected reversion times over extreme negative deviations raised from 2.49 to 4.68 (88%) and given positive deviations from 1.96 to 4.96 (more than 154%).

This *Euro effect* related with recovery times can be isolated considering the ratios $\frac{(\bar{y}_{ik} - \sigma_{y_{ik}})_{EM}}{(\bar{y}_{ik} - \sigma_{y_{ik}})}$ and $\frac{(\bar{y}_{ik} + \sigma_{y_{ik}})_{EM}}{(\bar{y}_{ik} + \sigma_{y_{ik}})}$ for the two periods - before and after the introduction of the Euro. For negative deviations, the ratio raised from 1.10 in 1962-1998 to 1.17 in 1999-2016 suggesting an increase in the regional imbalances given recessions and, more precisely, that the gap between the European countries after the introduction of the euro is higher inside than outside the Eurozone.

4.4 CONCLUSIONS

Our findings favor the hypothesis that the Euro generated a regime change in the macrodynamics of the economic space we consider and that this change impacted on the growth of the economies, imposing a process of divergence instead of convergence. For 1999-2016 we detect higher dispersion of the performances of the different economies than in the previous period, both for positive and negative deviations from the mean but also, in particular, we find that low growth rates correlate with high mean reversion time given negative deviations. High persistence or low speed mean reversion indicates divergence either through successive shocks or through endogenous economic changes. This can signify the presence of self-reinforcing mechanisms or political choices consistent with the formation of this new regime for the Euro period. As expected, the results of clusterization analysis for the period after 1999 confirm these results and we find the Southern countries of Europe forming a well defined cluster for that period, unlike in the past.

5 TIME INHOMOGENEOUS MULTIVARIATE MARKOV CHAINS: DETECTING AND TESTING MULTIPLE STRUCTURAL BREAKS OCCURRING AT UNKNOWN DATES

5.1 INTRODUCTION

Let $\{S_t, t = 0, 1, 2 \cdots, \infty\}$, hereinafter $\{S_t\}$, be a stochastic process that involves a sequence of discrete random variables with domain $E = \{1, \cdots, q\}$. Furthermore, $\{S_t\}$ is a first order Markov chain in the sense that:

$$P\left(S_t = i_0 \mid \mathcal{F}_{t-1}\right) = P\left(S_t = i_0 \mid S_{t-1} = i_1\right) \equiv P_{i_1 i_0}, \tag{41}$$

where $\mathcal{F}_{t-1}$ is the σ-field generated by all available information until the period $t-1$.

Given an initial condition and once the domain E is known, S_t can be fully characterized with the associated transition probability matrix (TPM) P. This matrix contains all possible one step ahead transitions generated in the space $E = \{1, \cdots, q\}$, such that, for the generic period t, may be written as:

$$\mathbf{P_t} = \begin{pmatrix} P_{t,11} & P_{t,12} & \cdots & P_{t,1q} \\ P_{t,21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ P_{t,q1} & \cdots & \cdots & P_{t,qq} \end{pmatrix} \tag{42}$$

Markov chain models have shown to be proficiently and interdisciplinary used. Notably in Economics [55], Finance [34, 84], Financial Markets [53, 64, 67], Biology [12, 37, 76], Environmental Sciences [79, 82, 96], Physics [14, 36], Linguistics[1][52], Medicine [48], Forecasting [26], Management [44], Sports [18], the estimation of expected hitting times [24, 66], Operational Research [4, 19, 95], Economic History [25], among others; see, e.g. [22, 81].

Notwithstanding the common denominator of this research depicts the assumption about the homogeneous nature of the Markov chain. In fact, in all of the aforementioned studies the Markov chain assumed to be homogeneous in the sense that $P_{t,i_1 i_0} = P_{i_1 i_0}$ and $\mathbf{P_t} = \mathbf{P}$. In other words, the transition probability matrix does not depend on time.

1 In 1913 A. A. Markov illustrated his chains for the first time with an example taken from literature. He investigated a sequence of 20,000 letters in Pushkin's text *Eugeny Onegin* to model probability transitions between consonants and vowels, [52]

Ignoring the nonhomogeneous nature of a stochastic process by disregarding the presence of structural breaks can lead to misleading conclusions.[2] As concluded by Hansen [41] *Structural change is pervasive in economic time series relationships, and it can be quite perilous to ignore. Inferences about economic relationships can go astray, forecasts can be inaccurate, and policy recommendations can be misleading or worse* [41, p. 237]. Structural breaks are, then again, a common issue in economical environments. In fact, the economic dynamics is characterised by deep complex, and mutating, patterns of interdependence between variables and aggregates. For this reason, it is of great relevance to investigate whether the quantities $P_{t,i_1 i_0}$ for $i_k \in E, k = 0, \cdots, q$ are time invariant or, in contrast, time dependent.

Even though several studies approached Markov chains, few studies approached the issue of non homogeneity. Among these we highlight here Tan and Yılmaz [91] and Polansky [72]. On the one hand, Tan and Yılmaz [91] used a likelihood ratio test to investigate homogeneity in Markov chains. The method can only test for one single break occurring at a known date. On the other hand, Polansky [72] presented a method to detect and estimate change-points in Markov chains. However, the limiting distribution of the test statistic is unknown so the p-values were computed through bootstrapping. Moreover, the method is restricted to first order univariate Markov chains.

In conclusion, these studies have globally the following limitations:

1. Only one break is allowed;
2. Or the break is assumed to occur at a known date or the limiting distribution is unknown (bootstrap techniques are mandatory);

This paper proposes a new methodology for estimating and testing while allowing for several structural breaks occurring at unknown dates in Markov chains. More precisely, this paper proposes a method to:

1. Estimate the breaks dates;
2. Test for structural breaks using nonstandard but known distributions, whereas bootstrap techniques are useless.

The rest of this article is organized as follows. Section 2 exposes our the theoretical framework, enumerating the assumptions that allow a VAR representation of a Markov chain and discussing the main results of inhomogeneity detection in a Markov chain. Section 3 presents a Monte Carlo simulation where the power and the size of the proposed method is analysed. Section 4 discusses some possible extensions of this methodology. Finally, Section 5 elaborates on the summary of the main results and concludes.

2 The consequences of ignoring a structural break are widely documented in the literature, namely in the unit root tests, see, e.g. Perron [70] and Rappoport and Reichlin [77]

5.2 THE MODEL AND ASSUMPTIONS

In this section, we present the basic econometric context upon which our analysis will be elaborated. Furthermore a new method for detecting and testing multiple structural breaks occuring at unknown dates in non homogeneous Markov chains is proposed. Our strategy, which consists of representing a Markov chain in the form of a VAR model[3], comprises three distinct phases, namely:

1. To identify the conditions under which a Markov chain admits a stable VAR representation;
2. To represent a Markov chain into a VAR form;
3. To propose a theoretical econometric framework within which it is possible to detect and test for multiple structural breaks occurring at unknown dates.

5.2.1 *A vectorial autoregressive representation of a stationary Markov chain*

Consider the following random q-dimensional vector

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} & \cdots & y_{kt} & \cdots & y_{qt} \end{pmatrix}' \tag{43}$$

whose k-th element y_{kt} equals $1\{S_t = k\}$, (here $1\{\cdot\}$ denotes the indicator function, such that $y_{kt} = 1$ if $S_t = k$ and 0 otherwise).

Moreover, when $S_t = i$ then k-th element of $\mathbf{y}_{t+1}$, $y_{k,t+1}$, is a r.v. such that

$$P(S_{t+1} = k \mid S_t = i) = P(y_{k,t+1} = 1 \mid y_{it} = 1) = P_{ik} \tag{44}$$

and, by Markovian property and, without any loss of generality, assuming a first order MC it follows that $\mathbf{E}[\mathbf{y}_{t+1} | S_t = i] = \mathbf{P}_{i\bullet}$, the i-th row of $\mathbf{P}$.

Given this result it follows that an ergodic MC with domain $E = \{1, \cdots, q\}$ admits the following system representation:

$$\begin{cases} y_{1t} = & P_{11}y_{1,t-1} + P_{21}y_{2,t-1} + \cdots + P_{q1}y_{q,t-1} + \varepsilon_{1t} \\ y_{2t} = & P_{12}y_{1,t-1} + P_{22}y_{2,t-1} + \cdots + P_{q2}y_{q,t-1} + \varepsilon_{2t} \\ \vdots & \vdots \\ y_{q-1,t} = & P_{1,q-1}y_{1,t-1} + P_{2,q-1}y_{2,t-1} + \cdots + P_{q,q-1}y_{q,t-1} + \varepsilon_{q-1,t}. \end{cases} \tag{45}$$

3 Markov chains have also proven to be a valuable tool when it comes to the approximation of VAR models, see e.g. Tauchen [92]. The estimated transition probabilities might be relevant to find numerical solutions to integral equations in the absence of integration.

Or, equivalently,

$$\begin{cases} y_{1t} = & z_t' \mathbf{P}_{\bullet 1} + \varepsilon_{1t} \\ \vdots & \vdots \\ y_{q-1,t} = & z_t' \mathbf{P}_{\bullet q-1} + \varepsilon_{q-1,t} \end{cases}, \tag{46}$$

where $z_t' = \begin{pmatrix} y_{1,t-1}, & \cdots & , y_{q,t-1} \end{pmatrix}$; $\varepsilon_{it} \equiv y_{it} - \mathbf{E}[y_{it} \mid \mathcal{F}_{t-1}]$ for $i = 1, \cdots, q-1$; and $\mathbf{P}_{\bullet k}$ is the k-th column of $\mathbf{P}$, $k = 1, \cdots q$.

Furthermore let the sufficient statistics for $\mathbf{P}$, n_{ik}, $\forall i, k \in E$, denote the number of transition frequencies of the type $i \to k$ in the sample, i.e.

$$n_{ik} = \sum_{t=1}^{n} 1\{S_t = k, S_{t-1} = i\}, \tag{47}$$

it can be shown that the likelihood function (the distribution is multinomial) is

$$l(P_{ik}) \propto \sum_i \sum_k n_{ik} \log(P_{ik}) \tag{48}$$

and the maximum likelihood estimator for P_{ik} is

$$\hat{P}_{ik} = \frac{\sum_{t=1}^{n} 1\{S_t = k, S_{t-1} = i\}}{\sum_{k=1}^{q} \sum_{t=1}^{n} 1\{S_t = k, S_{t-1} = i\}} = \frac{n_{ik}}{n_i} \tag{49}$$

see Basawa [11] and Billingsley [13].

It is useful to represent models (45) and (46) in the following matrix form:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{q-1t} \end{pmatrix} = \begin{pmatrix} z_t' & 0 & \cdots & 0 \\ 0 & z_t' & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & z_t' \end{pmatrix} \begin{pmatrix} \mathbf{P}_{\bullet 1} \\ \mathbf{P}_{\bullet 2} \\ \vdots \\ \mathbf{P}_{\bullet q-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{q-1t} \end{pmatrix} \tag{50}$$

or equivalently,

$$\begin{aligned} \underset{(q-1)\times 1}{\boldsymbol{y}_t} &= \left(\underset{(q-1)\times(q-1)}{\mathbf{I}} \otimes \underset{1\times q}{\boldsymbol{z}'} \right) \underset{q(q-1)\times q(q-1)}{\text{vec}(\mathbf{P}^*)} + \underset{(q-1)\times 1}{\boldsymbol{\varepsilon}_t} \\ &= \underset{(q-1)\times q(q-1)}{\boldsymbol{x}_t'} \times \underset{q(q-1)\times 1}{\mathbf{p}} + \underset{(q-1)\times 1}{\boldsymbol{\varepsilon}_t} \end{aligned} \tag{51}$$

where

$$\mathbf{P}^* = \begin{pmatrix} P_{11} & P_{12} & P_{1q-1} \\ P_{21} & P_{22} & P_{2q-1} \\ \vdots & \vdots & \vdots \\ P_{q1} & P_{q2} & P_{qq-1} \end{pmatrix} \tag{52}$$

$$= \begin{pmatrix} \mathbf{P}_{\bullet 1} & \mathbf{P}_{\bullet 2} & \cdots & \mathbf{P}_{\bullet q-1} \end{pmatrix}, \tag{53}$$

$$\mathbf{p} \equiv \text{vec}\left(\mathbf{P}^*\right) = \begin{pmatrix} \mathbf{P}_{\bullet 1} \\ \mathbf{P}_{\bullet 2} \\ \vdots \\ \mathbf{P}_{\bullet q-1} \end{pmatrix}, \tag{54}$$

and ε_t is a martingale difference sequence with covariance matrix $\mathbf{\Sigma} \equiv \mathbf{E}\left[\varepsilon_t \varepsilon_t'\right]$ given by

$$\mathbf{\Sigma} = \begin{pmatrix} \pi_1 - \sum_{k=1}^q \pi_k P_{k1}^2 & -\sum_{k=1}^q P_{k1} P_{k2} \pi_k & \cdots & -\sum_{k=1}^q P_{k1} P_{kq} \pi_k \\ -\sum_{k=1}^q P_{k1} P_{k2} \pi_k & \pi_2 - \sum_{k=1}^q \pi_k P_{k2}^2 & \cdots & -\sum_{k=1}^q P_{k2} P_{kq} \pi_k \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{k=1}^q P_{k1} P_{kq} \pi_k & -\sum_{k=1}^q P_{k2} P_{kq} \pi_k & & \pi_q - \sum_{k=1}^q \pi_k P_{kq}^2 \end{pmatrix}, \tag{55}$$

where π_k, $k = 1, \cdots, q$ denote the the stationary state k probability, see Lemma (1), Mathematical Appendix.

Expression (51) suggests that, under certain conditions, a Markov chain may assume a stable VAR representation.[4] A clear implication of this circumstance is that detecting, and testing, non homogeneities in a Markov chain stochastic process can be can be treated as a problem of testing for structural breaks in linear systems of equations.

In this sense, the final model may be written as VAR model (subject to m breaks) with $m+1$ distinct regimes, or segments, such that:

$$y_t = \begin{cases} x_t' \mathbf{p}_1 + \varepsilon_1, & \text{for } t = T_0 + 1, \cdots, T_1 \\ x_t' \mathbf{p}_2 + \varepsilon_2, & \text{for } t = T_1 + 1, \cdots, T_2 \\ \vdots & \vdots \\ x_t' \mathbf{p}_m + \varepsilon_m, & \text{for } t = T_{m-1} + 1, \cdots, T_m \\ x_t' \mathbf{p}_{m+1} + \varepsilon_{m+1}, & \text{for } t = T_m + 1, \cdots, T_{m+1} \end{cases} \tag{56}$$

4 We will explain such conditions later

or

$$y_t = x_t' p_j + \varepsilon_j \tag{57}$$

where ε_j is a martingale difference sequence with covariance matrix $\boldsymbol{\Sigma}_j$; $T_0 = 0$ and $T_{m+1} = T$; for $T_{j-1} + 1 \leqslant t \leqslant T_j, j = 1, \cdots, m+1$.

The main objective of this article is to consistently estimate the stacked vectors of parameters $\theta \equiv (\mathbf{p}_1, \cdots, \mathbf{p}_{m+1}; \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_{m+1})$, and the $m - dimensional$ break dates vector $\mathcal{T} = (T_1, \cdots, T_m)$.

With regard to estimating and testing for multiple structural breaks in systems of linear equations, Bai, Lumsdaine, and Stock [7] proposed a method for testing one single break and Hansen [42] considered multiple breaks occurring at known dates in a cointegrated system. Bai [6] considered a problem of testing multiple breaks occurring at unknown dates but, as far as we know, the most general theoretical framework for testing the presence of structural breaks occurring at unknown dates is the one proposed by Qu and Perron [74]. In truth, as noted by Perron [71], the latter method has several advantages over the former. Namely, the possibility of testing within- and cross-equation restrictions of the type $\mathbf{g}(\mathbf{p}^\star, \text{vec}(\boldsymbol{\Sigma}^\star)) = \mathbf{0}$, where $\mathbf{p}^\star \equiv \left(\mathbf{p}_1', \mathbf{p}_2', \cdots, \mathbf{p}_{m+1}'\right)'$ and $\boldsymbol{\Sigma}^\star \equiv (\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_{m+1})$. This is relevant in the sense that several interesting special cases can be dealt within the framework of Qu and Perron [74]: i) partial structural change models (only a subset of the parameters are subject to change), ii) block partial structural change models (only a subset of the equations are subject to change); iii) ordered break models (the breaks can occur in a particular order across subsets of equations); among others. Additionally, the problem of structural breaks in Markov chain models for panel data can be addressed. See Perron [71] for a discussion of some methodological issues related to estimation and testing of structural changes in the linear models.

5.2.2 *The model: main assumptions*

In this subsection we present the main econometric theory that supports our strategy. For this purpose, the following assumptions on the Markov chain are imposed.

Assumption A6. $\{S_t\}$ *is a first order, possibly,* m*-inhomogeneous Markov chain.*

Remark 1. *A Markov chain is said to be* m*-inhomogeneous if and only if the maximum number of distinctive transition probability matrices is* m *- the number of segments is* $m+1$*. Or, in other words, the number of breaks is* m*.*

Assumption A7. $\{S_t\}$ *is a positive recurrent MC and aperiodic in the sense that its states are positive recurrent aperiodic, in each potential segment* $j = 1, \cdots, m+1$*.*

Remark 2. *A positive recurrent and aperiodic MC is said to be ergodic or irreducible. In these circumstances the process* $\{S_t\}$ *admits a unique stationary distribution in each segment given by* $\mathbf{\Pi}_j$*, by the Perron-Frobenius theorem and given that each* $\mathbf{P}_t$*, for* $T_{j-1}+1 \leqslant t \leqslant T_j, j = 1, \cdots, m+1$ *matrix has an eigenvalue* λ_{jt} *equal to one and all of them have roots outside the unity circle, see Suhov and Kelbert [89].*

Assumption A8. $\exists \ell_0$: *the minimum eigenvalues of*

$$(1/\ell) \sum_{j=T_j^0+1}^{T_j^0+\ell} x_t x_t'$$

and of

$$(1/\ell) \sum_{j=T_j^0-\ell}^{T_j^0} x_t x_t'$$

are bounded away from zero, for $j = 1, \cdots, m; \forall \ell > \ell_0$.

Assumption A9. *The matrices* $\sum_{t=k}^{\ell} x_t x_t'$ *are invertible* $\forall\ \ell - k \geqslant k_0$*, for some* $0 < k_0 < \infty$.

Assumption A8 requires that there is no local perfect collinearity in the regressors near the break dates. This ensures that the break dates are identifiable. Assumption A9 is a standard invertibility condition.

These assumptions are plausible given that the Markov chain is positive recurrent and aperiodic for each segment.

Within our theoretical framework the following propositions arise.

Propositon 5. *Under Assumptions A6 and A7, the OLS estimator for (51) is, regardless of the sample size, numerically equal to the one obtained through the ML that assumes a multinomial distribution in (49)*

Proof. See Mathematical Appendix ∎

Henceforth we will use the superscript 0 to denote the true values of model parameters, such that $(\mathbf{p}_1^0, \cdots, \mathbf{p}_{m+1}^0)$, $\left(\mathbf{\Sigma}_1^0, \cdots, \mathbf{\Sigma}_{m+1}^0\right)$, and $\mathfrak{T}^0 = (T_1^0, \cdots, T_m^0)$ denotes, respectively, the true value of the mean equations parameters, the true values of the error covariance matrix and the true break dates. Furthermore let

$$\theta^0 \equiv \left(\mathbf{p}_1^0, \cdots, \mathbf{p}_{m+1}^0; \mathbf{\Sigma}_1^0, \cdots, \mathbf{\Sigma}_{m+1}^0\right) \tag{58}$$

represent the stacked value of the true parameters of the model.

Propositon 6. *Under Assumptions A6 and A7 we have*

$$\ell_j^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} x_t x_t' \overset{a.s.}{\to} Q_j, \tag{59}$$

a non random positive definite matrix, as $\ell_j \to \infty$*, for each* $j = 1, \cdots, m+1$ *and* $\ell_j \leqslant T_j^0 - T_{j-1}^0 + 1$.

Proof. See Mathematical Appendix ■

This proposition ensures the verification of all necessary conditions for the application of the central limit theorem.

Propositon 7. *Under Assumptions A6 and A7:*

1. $\{x_t \varepsilon_t, \mathcal{F}_t\}$ *forms a martingale difference sequence;*
2. $\mathbf{E}[x_t \varepsilon_t] = 0$*;*

Proof. See Mathematical Appendix ■

Proposition 7 naturally holds if x_t is replaced by ε_t (or by $\varepsilon_t \varepsilon_t' - \boldsymbol{\Sigma}_j$).

5.2.3 *The estimation procedure*

The main objective of this section is to discuss the estimation procedure, as well as to expose the limiting distribution of parameter estimators. Let us focus now on the estimation of the parameters θ, we will elaborate on the estimation and inference of the break fractions latter on. Assuming that the transition probabilities change in known periods[5], conditioning on a partition of the sample $\mathcal{T}$, the parameters of the model (57) can be consistently estimated through the quasi-maximum likelihood method. The quasi-likelihood function is:

$$L(\mathbf{p}, \boldsymbol{\Sigma}) = \prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f\left(y_t \mid x_t; \mathbf{p}_j, \boldsymbol{\Sigma}_j\right) \tag{60}$$

and the the quasi-likelihood ratio may be written as:

$$LR = \frac{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f\left(y_t \mid x_t; \mathbf{p}_j, \boldsymbol{\Sigma}_j\right)}{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}^0+1}^{T_j^0} f\left(y_t \mid x_t; \mathbf{p}_j^0, \boldsymbol{\Sigma}_j^0\right)} \tag{61}$$

with

$$f\left(y_t \mid x_t; \mathbf{p}_j, \boldsymbol{\Sigma}_j\right) = \frac{1}{(2\pi)^{(q-1)/2} \left|\Sigma_j\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(y_t - x_t' \mathbf{p}_j\right)' \boldsymbol{\Sigma}_j^{-1}\left(y_t - x_t' \mathbf{p}_j\right)\right\} \tag{62}$$

The estimators for $\mathbf{p}_j$ and $\boldsymbol{\Sigma}_j$ are obtained as

$$\left(\hat{\mathbf{p}}_j, \hat{\boldsymbol{\Sigma}}_j\right) \equiv \underset{(\mathcal{T}, \mathbf{p}_j, \boldsymbol{\Sigma}_j)}{\operatorname{argmax}} \log\left(LR(\mathbf{p}, \boldsymbol{\Sigma})\right) \tag{63}$$

5 Or that we have consistent estimators for $\mathcal{T}$.

resulting in the following joint closed solutions

$$\hat{\mathbf{p}}_j = \left(\sum_{t=T_{j-1}+1}^{T_j} x_t x_t' \right)^{-1} \sum_{t=T_{j-1}+1}^{T_j} x_t y_t \tag{64}$$

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{T_j - T_{j-1}} \sum_{t=T_{j-1}+1}^{T_j} \left(y_t - x_t' \hat{\mathbf{p}}_j\right) \left(y_t - x_t' \hat{\mathbf{p}}_j\right)' \tag{65}$$

with this maximization being taken over some set of admissible partitions $\mathfrak{T}$ in the set:

$$\boldsymbol{\Lambda}_\varepsilon = \left\{(T\lambda_1, ..., T\lambda_m)); \left|\lambda_{j+1} - \lambda_j\right| \geqslant \varepsilon, \lambda_1 \geqslant \varepsilon, \lambda_m \leqslant 1 - \varepsilon\right\}, \tag{66}$$

where ε is a trimming parameter that imposes a minimal length for each segment and λ_j denotes the break fractions in such a way that $T_j = T\lambda_j$.

The consistency and asymptotic normality of $\hat{\mathbf{p}}_j$ can easily be shown for the reason $\hat{\mathbf{p}}_j$ can be seen as an M-estimator. In fact, the associated objective function might be written as a sample average $T^{-1} \sum_{t=1}^{T} \mathbf{m}(\boldsymbol{w}, \theta)$. Furthermore, the log likelihood ratio (61) can be decomposed as follows:

$$\begin{aligned}
\log LR &= \log \left(\frac{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f\left(y_t \mid x_t; \mathbf{p}_j, \boldsymbol{\Sigma}_j\right)}{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}^0+1}^{T_j^0} f\left(y_t \mid x_t; \mathbf{p}_j^0, \boldsymbol{\Sigma}_j^0\right)} \right) \\
&= \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} \log f\left(y_t \mid x_t; \mathbf{p}_j, \boldsymbol{\Sigma}_j\right) - \\
&\quad \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \log f\left(y_t \mid x_t; \mathbf{p}_j^0, \boldsymbol{\Sigma}_j^0\right) \\
&= \sum_{t=1}^{T} \left[\left\{ 1\{t \in seg_1\} l(\boldsymbol{w}_t; \theta_1) - 1\left\{t \in seg_1^0\right\} l\left(\boldsymbol{w}_t; \theta_1^0\right) \right\} + \cdots + \right. \\
&+ \left. \left\{ 1\{t \in seg_{m+1}\} l(\boldsymbol{w}_t; \theta_{m+1}) - 1\left\{t \in seg_{m+1}^0\right\} l\left(\boldsymbol{w}_t; \theta_{m+1}^0\right) \right\} \right] \\
&= \sum_{t=1}^{T} \left[\sum_{j=1}^{m+1} \left\{ 1\left\{t \in seg_j\right\} l\left(\boldsymbol{w}_t; \theta_j\right) - 1\left\{t \in seg_j^0\right\} l\left(\boldsymbol{w}_t; \theta_j^0\right) \right\} \right] \\
&= \sum_{t=1}^{T} \mathbf{m}(\boldsymbol{w}_t, \theta)
\end{aligned} \tag{67}$$

where $l\left(\mathbf{w}_t; \theta_j\right) \equiv \log f\left(\mathbf{y}_t \mid \mathbf{x}_t; \mathbf{p}_j, \boldsymbol{\Sigma}_j\right)$, and $1\left\{t \in seg_j\right\}$ means that we are in the j-th segment or, in other words, that $T_{j-1} + 1 \leqslant t \leqslant T_j$ for $j = 1, \cdots, m$, and maximizing the objective function $\sum_{t=1}^{T} \mathbf{m}(\boldsymbol{w}, \theta)$ is equivalent to maximizing $T^{-1} \sum_{t=1}^{T} \mathbf{m}(\boldsymbol{w}, \theta)$.

As the parameter set is compact (θ involves only the transition probabilities P_{ij} which are obviously bounded between 0 and 1) and the identification condition automatically holds by Proposition 6, we just need to assume the standard dominance condition

$$\mathbf{E}\left[\sup_{\theta\in\Theta}\left|l\left(w_t,\theta_j\right)\right|\right] < \infty, j = 1,\cdots,m+1$$

and Newey and McFadden [62] Theorem 2.5 is verified, hence we have $\hat{\mathbf{p}}_j \xrightarrow{p} \mathbf{p}_j^0$.

Moreover, assuming that for each $j = 1,\cdots,m$, the standard mild conditions for the asymptotic normality of an M-Estimator[6] and Newey and McFadden [62] Theorem 3.3 is automatically verified and we have, in these circumstances,

$$\sqrt{T_j - T_{j-1}}\left(\hat{\mathbf{p}}_j - \mathbf{p}^0\right) \xrightarrow{d} N\left(0, Avar\left(\hat{\mathbf{p}}_j\right)\right), \tag{68}$$

where $Avar\left(\hat{\mathbf{p}}_j\right) = \left[\text{plim}\left(\frac{1}{T_j-T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j} z_t z_t'\right)^{-1} \otimes \mathbf{\Sigma}_j\right]$, for $j = 1,\cdots,m$.

Let us know focus on the estimation of the break dates

$$\left(\hat{T}_1,\cdots,\hat{T}_m\right) = \left(T\hat{\lambda}_1,...,T\hat{\lambda}_m\right)$$

elaborating on the estimation of the parameters in a very general setup such that (within and cross-equation) restrictions of the type $\mathbf{g}\left(\mathbf{p}^\star, \text{vec}\left(\mathbf{\Sigma}^\star\right)\right) = 0$ are allowed[7]. We will follow the Qu and Perron [74] strategy adapted to our model.

To establish theoretical results about the consistency and limit distribution of the estimates of the break dates, some standard conditions on the asymptotic framework and on the break dates must be adopted. We also assume here that the break dates are asymptotically distinct (A10); and some conditions under which the breaks are asymptotic nonnegletable (A11). More precisely we consider the following Assumptions.

Assumption A10. *The following inequalities hold* $0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1$ *and* $T_i^0 = \left[T\lambda_i^0\right]$, *where* $[\cdot]$ *denotes the greatest integer for* $i = 0,\cdots,m+1$.

Assumption A11. *The magnitudes of the shifts satisfy* $\Delta\mathbf{p}_j^0 = \mathbf{p}_j^0 - \mathbf{p}_{j-1}^0 = \nu\delta_j$, $\Delta\mathbf{\Sigma}_j^0 = \mathbf{\Sigma}_j^0 - \mathbf{\Sigma}_{j-1}^0 = \nu\mathbf{\Phi}_j$, *for* $\left(\delta_j, \mathbf{\Phi}_j\right) \neq 0$, *and do not depend on* T. ν *is a positive quantity independent of* T; *or we have* $\nu \to 0$ *but* $\sqrt{T}\nu/\left(\log T\right)^2 \to \infty$, *being* ν *is a sequence of positive numbers.*

6 More precisely, that θ_0 is in the interior of Θ; $f\left(y_t \mid x_t;\theta_j\right)$ is twice continuously differentiable in θ_j for all $\left(y_t, x_t\right)$; $\mathbf{E}\left[\mathbf{s}\left(w_t,\theta_0\right)\right] = 0$ and $-\mathbf{E}\left[\mathbf{H}\left(w_t,\theta_0\right)\right] = \mathbf{E}\left[\mathbf{s}\left(w_t,\theta_0\right)\mathbf{s}\left(w_t,\theta_0\right)'\right]$ is non singular; and local dominance condition for the Hessian; where $\mathbf{s}\left(w_t,\theta\right) = \frac{\partial m(w_t,\theta)}{\partial\theta}$ and $\mathbf{H}\left(w_t,\theta_0\right) = \frac{\partial s(w_t,\theta)}{\partial\theta'}$

7 Those restrictions might play an important role because, among other cases, they permit the generation of partial change models (only a subset of coefficients are allowed to change). This class of models allows inhomogeneities to be tested in just a few lines of the matrix of transition probabilities, while the rest remain homogeneous. Furthermore, the imposition of relevant restrictions in the parameters of the model might also contribute to power increments.

Assumption A11 states that the magnitudes of the shifts can be either fixed (ν is a positive number independent of T) capturing large shifts asymptotically or shrinking (ν shrinks and $\sqrt{T}\nu/(\log T)^2 \to \infty$) corresponding to small shifts in finite samples. It is worth noting that the assumption on the nature of the magnitudes does not impact the test itself or its asymptotic distribution, as we will show later.[8] Research by Perron [71] discusses in length the implications of considering fixed shifts or shrinking shifts.

As noted by Qu and Perron [74] these Assumptions ensure four important results regarding the estimation process of the break dates. First, T_j being unknown does not change the distribution of the estimators. Formally, under Assumptions A6 to A11 the limiting distribution of $\sqrt{T}\left(\hat{\mathbf{p}}_j - \mathbf{p}_j^0\right)$ is the same regardless of whether the breaks occur at known or at unknown dates, hence the asymptotic normality and consistency of the estimator of the probabilities is established.

Second, the convergence rates of the estimators are as follows:

$$\begin{aligned} T\nu^2\left(\hat{T}_j - T^0\right) &= O_p(1), \text{ for } j = 1, \cdots, m && (69)\\ \sqrt{T}\left(\hat{\mathbf{p}}_j - \mathbf{p}_j^0\right) &= O_p(1), \text{ for } j = 1, \cdots, m+1 && (70)\\ \sqrt{T}\left(\hat{\mathbf{\Sigma}}_j - \mathbf{\Sigma}_j^0\right) &= O_p(1), \text{ for } j = 1, \cdots, m+1 && (71) \end{aligned}$$

which means that the break dates (or the break fractions) converge faster than $\hat{\mathbf{p}}_j$, such that the asymptotic distribution of the latter is not affected by the former.

Third. As a consequence of the last result, that the maximazation of the likelihood function might be done in a subset of the parameter set C_M and in a neighbourhood of the respective true values, such that:

$$\begin{aligned} C_M = & \left\{\left(\mathcal{T}, \mathbf{p}_j, \mathbf{\Sigma}_j\right) : \nu^2 \left|T_j - T_j^0\right| \leqslant M \text{ for } j = 1, \cdots, m, \right. && (72)\\ & \left. \left|\sqrt{T}\left(\mathbf{p}_j - \mathbf{p}_j^0\right)\right| \leqslant M, \left|\sqrt{T}\left(\mathbf{\Sigma}_j - \mathbf{\Sigma}_j^0\right)\right| \leqslant M \text{ for } j = 1, \cdots, m+1\right\} \end{aligned}$$

where M, which can be set to be arbitrarily large, denotes the maximum number of breaks allowed.

Finally, the log-likelihood ratio might be decomposed into two asymptotically independent components - one that concerns the estimation of the break dates and another one that refers to the estimation of the stacked vector of parameters $\boldsymbol{\theta}$. Let lr and rlr denote, respectively, the log likelihood ratio and the restricted likelihood ratio, such that the objective function is:

$$rlr = lr + \boldsymbol{\lambda}'\mathbf{g}\left(\mathbf{p}^\star, \text{vec}\left(\mathbf{\Sigma}^\star\right)\right) \quad (73)$$

Under Assumptions A6 to A11, rlr may be decomposed as follows:

8 On the contrary, this assumption only influences some minor technical issues related to the distribution of the break dates.

$$\begin{aligned}
\max_{(\mathcal{T},\mathbf{p},\boldsymbol{\Sigma})\in C_M} \mathrm{rlr} &= \max_{\mathcal{T}\in C_M,\mathbf{p}^0,\boldsymbol{\Sigma}^0} \sum_{j=1}^{m} \mathrm{lr}_j^{(1)}\left(T_j - T_j^0\right) & (74)\\
&+ \max_{(\mathbf{p},\boldsymbol{\Sigma})\in C_M,\mathcal{T}^0} \left[\sum_{j=1}^{m+1} \mathrm{lr}_j^{(2)} + \boldsymbol{\lambda}'\mathbf{g}\left(\mathbf{p},\boldsymbol{\Sigma}\right)\right] & (75)\\
&+ o_p\left(1\right)
\end{aligned}$$

where,

$$\begin{aligned}
\mathrm{lr}_j^{(1)}(0) &= 0,\\
\mathrm{lr}_j^{(1)}(r) &= \frac{1}{2}\sum_{t=T_j^0+r}^{T_j^0} \boldsymbol{\varepsilon}_t'\left[\left(\boldsymbol{\Sigma}_j^0\right)^{-1} - \left(\boldsymbol{\Sigma}_{j+1}^0\right)^{-1}\right]\boldsymbol{\varepsilon}_t - \frac{r}{2}\left(\log\left|\boldsymbol{\Sigma}_j^0\right| - \log\left|\boldsymbol{\Sigma}_{j+1}^0\right|\right)\\
&- \frac{1}{2}\sum_{t=T_j^0+r}^{T_j^0} \left(\mathbf{p}_j^0 - \mathbf{p}_{j+1}^0\right)'\mathbf{x}_t\left(\boldsymbol{\Sigma}_{j+1}^0\right)^{-1}\mathbf{x}_t'\left(\mathbf{p}_j^0 - \mathbf{p}_{j+1}^0\right)\\
&- \sum_{t=T_j^0+r}^{T_j^0} \left(\mathbf{p}_j^0 - \mathbf{p}_{j+1}^0\right)'\mathbf{x}_t\left(\boldsymbol{\Sigma}_{j+1}^0\right)^{-1}\boldsymbol{\varepsilon}_t
\end{aligned}$$

for $r = -1, -2, \cdots,$

$$\begin{aligned}
\mathrm{lr}_j^{(1)}(r) &= -\frac{1}{2}\sum_{t=T_j^0+1}^{T_j^0+r} \boldsymbol{\varepsilon}_t'\left[\left(\boldsymbol{\Sigma}_j^0\right)^{-1} - \left(\boldsymbol{\Sigma}_{j+1}^0\right)^{-1}\right]\boldsymbol{\varepsilon}_t - \frac{r}{2}\left(\log\left|\boldsymbol{\Sigma}_j^0\right| - \log\left|\boldsymbol{\Sigma}_{j+1}^0\right|\right)\\
&- \frac{1}{2}\sum_{t=T_j^0+1}^{T_j^0+r} \left(\mathbf{p}_j^0 - \mathbf{p}_{j+1}^0\right)'\mathbf{x}_t\left(\boldsymbol{\Sigma}_{j+1}^0\right)^{-1}\mathbf{x}_t'\left(\mathbf{p}_j^0 - \mathbf{p}_{j+1}^0\right)\\
&- \sum_{t=T_j^0+1}^{T_j^0+r} \left(\mathbf{p}_j^0 - \mathbf{p}_{j+1}^0\right)'\mathbf{x}_t\left(\boldsymbol{\Sigma}_{j+1}^0\right)^{-1}\boldsymbol{\varepsilon}_t
\end{aligned}$$

for $r = 1, 2, \cdots,$ and

$$\begin{aligned}
\mathrm{lr}_j^{(2)} &= \frac{1}{2}\sum_{t=T_j^0+1}^{T_j^0} \left(\mathbf{y}_t - \mathbf{x}_t'\mathbf{p}_j\right)'\boldsymbol{\Sigma}_j^{-1}\left(\mathbf{y}_t - \mathbf{x}_t'\mathbf{p}_j\right) - \frac{T_j^0 - T_{j-1}^0}{2}\log\left|\boldsymbol{\Sigma}_j\right|\\
&+ \frac{1}{2}\sum_{t=T_j^0+1}^{T_j^0} \left(\mathbf{y}_t - \mathbf{x}_t'\mathbf{p}_j\right)'\left(\boldsymbol{\Sigma}_j^0\right)^{-1}\left(\mathbf{y}_t - \mathbf{x}_t'\mathbf{p}_j\right) + \frac{T_j^0 - T_{j-1}^0}{2}\log\left|\boldsymbol{\Sigma}_j^0\right|.
\end{aligned}$$

This result implies that the estimator of the break dates are asymptotic independent of the estimator of θ. Additionally, eventual restrictions on the parameters do not affect the distribution of the break dates. Moreover, the estimation procedure is bietapic. Firstly, the break dates are consistently estimated assuming that we know the true values of the parameters θ^0, then the mean parameters are estimated, possibly subject to restrictions of the type $g(\mathbf{p}^\star, \text{vec}(\mathbf{\Sigma}^\star)) = 0$, keeping the break dates fixed at their true values $\mathcal{T}^0$. Thus, under fixed magnitudes of shifts, it is straightforward to derive the asymptotic distribution of the estimates on the break dates.[9]

Propositon 8. *Under Assumptions A6 to A11, assuming a fixed ν, we have:*

$$\hat{T}_j - T_j^0 \xrightarrow{d} \textit{argmax}_r \, lr_j^{(1)}(r) \text{ for } j = 1, \cdots, m \tag{76}$$

Proof. See Qu and Perron [74]. ∎

It is worth noting some considerations about the computational procedure underlying the estimation of the model parameters. A standard grid search procedure would require the computation of a number of QMLE of an order of magnitude of T^m, which would be virtually impossible with $m > 2$. However, the maximum number of possible segments is actually $\sum_{t=1}^{T} t = 1 + 2 + \cdots + T = \frac{T(1+T)}{2}$ and therefore of an order of magnitude of T^2, no matter the number of breaks m. Thus, a method is required to select which combination of segments leads to a minimum value of the objective function. Here, we follow the dynamic programming algorithm, based on a iterative GLS approach to evaluate the likelihood functions for all segments, proposed by Bai and Perron [9] and extended by Qu and Perron [74, pp 476-478].

5.2.4 *Testing for multiple and endogenous inhomogeneities in Markov chains*

In this section we consider the problem of testing for inhomogeneities in a Markov chain. Put otherwise, testing structural changes in the one-step transition probabilities occurring at unknown periods. Without any lost of generality we assume a pure structural change model, such that all parameters are allowed to vary over time. In a first moment we will expose the standard likelihood ratio test. This statistic intends to determine a presence of at least one structural break. Next, we discuss the potential of two possible extensions concerning confirmatory analysis. A sequential test that allows us to select the number of changes, given that, sequentially, we test a null of l breaks against $l+1$ breaks - the $Seq(\ell+1|\ell)$; and the $WDmax$ for testing no breaks against an unknown (up to some pre-specified maximum) number of breaks. These procedures are interesting because they do not require

9 For the derivation of the asymptotic distribution of the break dates estimator under shrinking magnitudes of shifts see, e.g. Qu and Perron [74, pp 471-472] or Bai [6, p312]

the prespecification of the number of breaks under the alternative hypothesis, unlike the sup LR test.

Nevertheless, the following two Assumptions, on the regressors and on the errors, must be adopted to obtain theoretical results about limiting distribution of the tests, under the null hypothesis of no breaks, $m = 0$.

Assumption A12. *We have* $T^{-1}\sum_{t=1}^{[Ts]} x_t x_t' \xrightarrow{p} sQ$ *uniformly in* $s \in [0,1]$ *for some* Q *positive definite.*

Assumption A12 imposes that the limit moment matrix of the regressors is homogeneous over all sample.

Assumption A13. *We have* $\mathbf{E}\left[\varepsilon_t \varepsilon_t'\right] = \mathbf{\Sigma}^0\ \forall t$; $T^{-1/2}\sum_{t=1}^{[Ts]} x_t \varepsilon_t \xrightarrow{d} \mathbf{\Phi}^{1/2}\boldsymbol{W}(s)$ *where* $\boldsymbol{W}(s)$ *is a vector of independent Wiener processes, and*
$\mathbf{\Phi} = \operatorname{plim} T^{-1}\sum_{t=1}^{T} x_t \left(I_{(q-1)} \otimes \mathbf{\Sigma}^0\right) x_t'$.

Also we have, $T^{-1/2}\sum_{t=1}^{[Ts]}\left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - \mathbf{I}_{(q-1)}\right) \xrightarrow{d} \boldsymbol{\xi}(s)$, *where* $\boldsymbol{\xi}(s)$ *is a* $(q-1)$ *square matrix of Brownian motion processes with* $\mathbf{\Omega} = var\left(vec\left(\boldsymbol{\xi}(1)\right)\right)$; *and* $\boldsymbol{\eta}_t \equiv \left(\eta_{t1}, \cdots, \eta_{t,q-1}\right)' = \left(\mathbf{\Sigma}^0\right)^{-1/2}\varepsilon_t$. *In addition we assume that* $\mathbf{E}\left[\eta_{tk}\eta_{tl}\eta_{tm}\right] = 0\ \forall k, l, m, t; k \neq m \vee k \neq l$.

Assumption A13 states that the functional central limit theorem might be employed. This is a mild weak condition, given that, by construction, under null the hypothesis of homogeneity, $x_t \varepsilon_t$ is a martingale difference sequence with respect to $\mathcal{F}_t$. Moreover, x_t and ε_t are bounded which ensures the existence of all moments of $x_t \varepsilon_t$, in particular, the Corollary 29.19 of Davidson [27] can be immediately applied. Both Assumption A12 and A13 are crucial to the derivation of the limiting distribution of the tests under H_0.

Regarding the limiting distribution of the sup $LR(m, q, \varepsilon)$ test, let

$$\tilde{\mathbf{p}} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t \tag{77}$$

denote the transition probabilities estimator under the null hypothesis of homogeneity (absence of structural breaks) and let $\widetilde{L}$ denote the associated likelihood, where

$$\log \widetilde{L} = -\frac{T(q-1)}{2}\left(\log 2\pi + 1\right) - \frac{T}{2}\log\left|\tilde{\mathbf{\Sigma}}\right|, \tag{78}$$

being $\tilde{\mathbf{\Sigma}}$ the error covariance matrix under homogeneity, verifying

$$\tilde{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\tilde{\mathbf{p}}\right)\left(y_t - x_t'\tilde{\mathbf{p}}\right)'. \tag{79}$$

Let additionaly $\log \hat{L}\left(T_{1,} \cdots, T_{m,}\right)$ denote the log likelihood function for a given partition $\mathcal{T}$ associated with the QML (60).

The observable value for this test is the supremum of the likelihood ratio:

$$2\left[\log \hat{L}\left(T_{1,}\cdots,T_{m,}\right)-\log \widetilde{L}\right] \tag{80}$$

evaluated over all possible partitions Λ_ε (expression 66). Formally, we have

$$\begin{aligned}\sup LR(m,q,\varepsilon) &= \sup_{(\lambda_1,\cdots,\lambda_m)\in\boldsymbol{\Lambda}_\varepsilon} 2\left[\log \hat{L}\left(T_1,\cdots,T_m\right)-\log \widetilde{L}\right]\\ &= 2\left[\log \hat{L}\left(\hat{T}_1,\cdots,\hat{T}_m\right)-\log \widetilde{L}\right],\end{aligned} \tag{81}$$

where $\hat{T}_1,\cdots,\hat{T}_m$ results from the maximization (74). This statistic depends on three parameters: i) the number of breaks allowed, m ; ii) the trimming parameter, ε; and iii) the dimension of the space state of the Markov chain (the number of states). The critical values are presented in Bai and Perron [8, 9]. Regarding the limiting distribution of the $\sup LR(m,q,\varepsilon)$ statistic under the null hypothesis of homogeneity, Proposition 9 holds.

Propositon 9. *Under Assumptions A6 to A13 the limiting distirbution of the* $\sup LR$ *statistic is as follows.*

$$\sup LR \xrightarrow{d} \sup \sum_{j=1}^{m} \frac{\left\|\lambda_j W_q\left(\lambda_{j+1}\right)-\lambda_{j+1} W_q\left(\lambda_{j+1}\right)\right\|^2}{\left(\lambda_{j+1}-\lambda_j\right)\lambda_j\lambda_{j+1}} \tag{82}$$

where $W_q(\cdot)$ *denote a* q*-dimensional vector of independent Wiener processes.*

Proof. See Qu and Perron [74, p. 487]. ■

We extend our analysis to a sequential test procedure. Actually, the $Seq\left(\ell+1|\ell\right)$ statistic proposed by Bai and Perron [8] to the univariate case and adapted by Qu and Perron [74] to the multivariate case, can be used to select the number of different segments in an inhomogeneous Markov chain. Let us consider a model with ℓ breaks, whose estimates, $(\hat{T}_1,\cdots,\hat{T}_\ell)$, were obtained by a global maximization procedure. This statistic tests, sequentially, $H_0:\ell$ breaks against $H_1:\ell+1$ breaks by performing a single-break test for for each one of the segments $(\hat{T}_1,\cdots,\hat{T}_\ell)$ and then evaluating the significance of the maximum of the tests. Formally we have:

$$\begin{aligned}&Seq\left(\ell+1\mid\ell\right)=\\&\quad\max_{t\leqslant j\leqslant \ell+1}\sup_{\tau\in\boldsymbol{\Lambda}_{j,\varepsilon}} lr\left(\hat{T}_1,\cdots,\hat{T}_{j-1},\tau,\hat{T}_j,\cdots\hat{T}_\ell\right)-lr\left(\hat{T}_1,\cdots,\hat{T}_\ell\right),\end{aligned} \tag{83}$$

where

$$\boldsymbol{\Lambda}_{j,\varepsilon}=\left\{\hat{T}_{j-1}+\left(\hat{T}_j-\hat{T}_{j-1}\right)\varepsilon\leqslant\tau\leqslant\hat{T}_j-\left(\hat{T}_j-\hat{T}_{j-1}\right)\varepsilon\right\}. \tag{84}$$

The limiting distribution of this statistic under the null of ℓ breaks can be found in Qu and Perron [74] and the critical values were tabulated in Bai and Perron [8, 9].

An important feature concerning the $\sup \mathrm{LR}(m, q, \varepsilon)$ statistic relates to the need to specify a priori the number of breaks to be tested, m, under the alternative hypothesis. That is frequently not the case and one may still not want to specify the number of breaks under the alternative hypothesis. For this purpose, Bai and Perron [8] suggested a class of tests called *double maximum tests*. One of these statistics is the $WDmax\mathrm{LR}(M)$, that tests for a null of no breaks against an unknown number of breaks between 1 and some upper limit M. The mechanic of this test consists on the evaluation of the maximum of the supremum of the $\sup \mathrm{LR}(m, q, \varepsilon)$ over the number of possible breaks from 1 to M, as follows

$$WDmax\mathrm{LR}(m, q, \varepsilon)\,(M, q) = \max_{1 \leqslant m \leqslant M} a_m \times \sup_{(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon} \mathrm{LR}(m, q, \varepsilon). \quad (85)$$

The terms a_m are the weights of the test,[10]

$$a_m \equiv \frac{c\,(q, \alpha, 1)}{c\,(q, \alpha, m)}, \quad (86)$$

the ratio between the asymptotic critical values of the $\sup \mathrm{LR}(m, q, \varepsilon)$ for $m = 1$ $(c\,(q, \alpha, 1))$ and for $m = 1, \cdots, M$ $(c\,(q, \alpha, m))$, so that $a_1 = 1$. Critical values might be found in Bai and Perron [8, 10].

5.3 MONTE CARLO EXPERIMENTS

In this section we evaluate the size the the power of the tests through a Monte Carlo experiment. We consider a simple process with two categories S_t with three states ($q = 3$). The Markov chain is simulated, using the GAUSS package, according to the algorithm:

1. Define $m+1$ q-dimensional transition probability matrices whose elements are the probabilities
$$P_{j,i_1 i_0} \equiv P_j\,(S_t = i_0 | \, S_{t-1} = i_1) \quad (87)$$
(see the definition of the DGPs below);
2. Initialize the process $\{S_t\}$ by assigning values to $S_{T_{j-1}}$, for $j = 1, \cdots, m$ accordingly to the stationary distributions $\mathbf{\Pi}_j$;
3. Simulate a continuous random variable W uniformly distributed, $W \sim U\,(0, 1)$;
4. Given the initial values $S_{T_{j-1}}$ (step 2), simulate the process $\{S_t\}$, $t = T_{j-1} + 1, .., T_j$, for $j = 1, \cdots, m+1$ as follows:

10 An unweighted version of this test, the $UDmax\mathrm{LR}(m, q, \varepsilon)$, reported poor power, mainly as m increases [7, 10].

a) Let $P_{i,j} \equiv P_j\left(S_t = i | S_{1t-1} = i_1\right), t = T_{j-1}+1, .., T_j; j = 1, \cdots, m+1$;

b) $S_t = \begin{cases} 1 & \text{if} \quad 0 \leqslant W < P_{1,j} \\ 2 & \text{if} \quad P_{1,j} \leqslant W < P_{1,j} + P_{2,j} \\ 3 & \text{if} \quad P_{1,j} + P_{2,j} \leqslant W < 1; \end{cases}$

5. Repeat the steps 1-4 until $t = T_{m+1}$.

In all simulations, computed with 5000 replications, a trimming parameter is set to be $\varepsilon = 0.15$. Sample sizes of 250, 500 and 1000 were considered. The proportion of times when the null hypothesis was rejected is reported in Figures 13 to 24 (Appendix 2). Several levels of persistency of the regimes of the Markov chain are analysed. The analysis is also extended to the situation where the regimes are independent.

5.3.1 *Size Analysis*

We now examine three different DGP (i.e. three different cases) for homogeneous Markov chains, $m = 1$, so that $T_0 = 1$ and $T_1 = T$ as follows. In Case 1 the regimes of the process are generated without any persistency, while in Case 2 one can see highly persistent regimes. Finally, Case 3 depicts the situation were the regimes are independent.

Case 1:

$$\begin{aligned} P_{11} &= 0.4, P_{12} = 0.3, P_{13} = 0.3, \\ P_{21} &= 0.2, P_{22} = 0.4, P_{23} = 0.4, \\ P_{31} &= 0.6, P_{32} = 0.2, P_{33} = 0.2. \end{aligned}$$

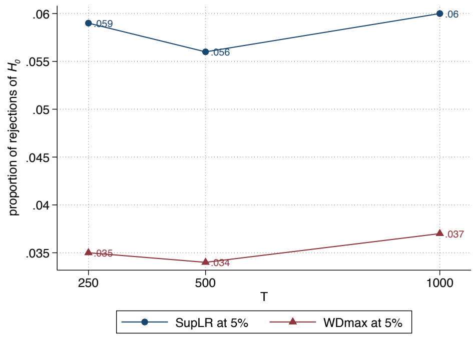
Case 2:

$$\begin{aligned} P_{11} &= 0.55, P_{12} = 0.25, P_{13} = 0.2, \\ P_{21} &= 0.25, P_{22} = 0.55, P_{23} = 0.2, \\ P_{31} &= 0.2, P_{32} = 0.25, P_{33} = 0.55. \end{aligned}$$

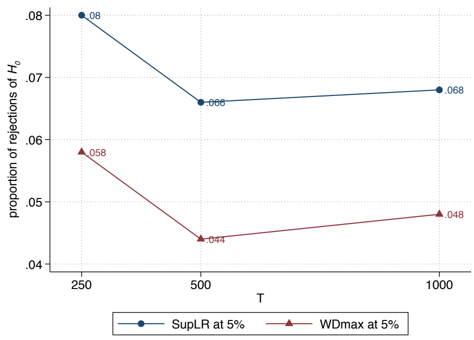
Case 3:

$$\begin{aligned} P_{11} &= 0.5, P_{12} = 0.3, P_{13} = 0.2, \\ P_{21} &= 0.5, P_{22} = 0.3, P_{23} = 0.2, \\ P_{31} &= 0.5, P_{32} = 0.3, P_{33} = 0.2. \end{aligned}$$

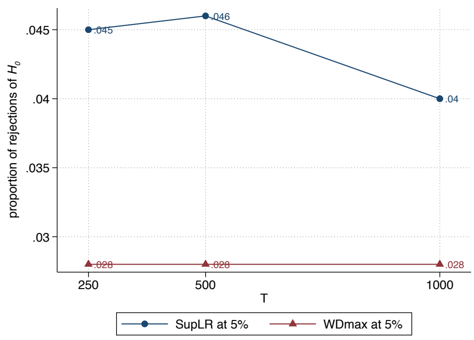
Next, we consider two cases: (1) one break is allowed, (2) two breaks are allowed. In the first case both the $\sup LR$ and the $WDmax$ tests 0 breaks against 1 break. To simplify the notation we consider $\sup LR(0 \mid 1)$ and $WDmax(0 \mid 1)$. In the second situation the $\sup LR$ tests 0 breaks against 2, $\sup LR(0 \mid 1)$, and the $WDmax$ 0 against an undetermined number of breaks up to 2, $WDmax(0 \mid 1, 2)$. We also consider results for the three standard significance levels, 1%, 5%, and 10%. Notwithstanding, we report here results for 5%, while the rest can be found in the Appendix section. Figures 13 and 14 display the results. In general, there appear to be no size distortions of the tests, as expected. The exception is the sup $LR(0 \mid 2)$ for the Case 2, where the regimes of the Markov exhibit some persistency, that tends to be slightly oversized (Figure 14b). It should also be noted that $WDmax$ tends to present an undersize behaviour in all situations, in that it tends to under-reject the null hypothesis.



(a) Case 1



(b) Case 2



(c) Case 3

Figure 13: Homogeneous Markov chain, one break is allowed (nominal size: 5%)

5.3.2 *Power Analysis*

To assess the power of the tests either in finite samples and asymptotically, our analysis includes six cases for inhomogeneous Markov chains. Cases 4, 5, 6 correspond to Markov chains generated with one structural break $m = 1$, so that $T_0 = 1$, $T_1 = [T/2]$ and $T_2 = T$. Cases 7, 8, 9 concern chains generated subject to three breaks ($T_0 = 1$, $T_1 = [T/3]$, $T_2 = [2T/3]$ and $T_3 = T$).

Case 4 (regime 1):

$$\begin{aligned} P_{1,11} &= 0.40, P_{1,12} = 0.30, P_{1,13} = 0.30, \\ P_{1,21} &= 0.20, P_{1,22} = 0.40, P_{1,23} = 0.40, \\ P_{1,31} &= 0.60, P_{1,32} = 0.20, P_{1,33} = 0.20. \end{aligned}$$

Case 4 (regime 2):

$$\begin{aligned} P_{2,11} &= 0.20, P_{2,12} = 0.40, P_{2,13} = 0.40, \\ P_{2,21} &= 0.20, P_{2,22} = 0.60, P_{2,23} = 0.60, \\ P_{2,31} &= 0.30, P_{2,32} = 0.40, P_{2,33} = 0.40. \end{aligned}$$

Case 5 (regime 1):

$$\begin{aligned} P_{1,11} &= 0.55, P_{1,12} = 0.25, P_{1,13} = 0.20, \\ P_{1,21} &= 0.25, P_{1,22} = 0.55, P_{1,23} = 0.20, \\ P_{1,31} &= 0.20, P_{1,32} = 0.25, P_{1,33} = 0.55. \end{aligned}$$

Case 5 (regime 2):

$$\begin{aligned} P_{2,11} &= 0.45, P_{2,12} = 0.20, P_{2,13} = 0.35, \\ P_{2,21} &= 0.30, P_{2,22} = 0.45, P_{2,23} = 0.25, \\ P_{2,31} &= 0.15, P_{2,32} = 0.40, P_{2,33} = 0.45. \end{aligned}$$

Case 6 (regime 1):

$$\begin{aligned} P_{1,11} &= 0.50, P_{1,12} = 0.30, P_{1,13} = 0.20, \\ P_{1,21} &= 0.50, P_{1,22} = 0.30, P_{1,23} = 0.20, \\ P_{1,31} &= 0.50, P_{1,32} = 0.30, P_{1,33} = 0.20. \end{aligned}$$

Case 6 (regime 2):

$$\begin{aligned} P_{2,11} &= 0.30, P_{2,12} = 0.40, P_{2,13} = 0.30, \\ P_{2,21} &= 0.30, P_{2,22} = 0.40, P_{2,23} = 0.30, \\ P_{2,31} &= 0.30, P_{2,32} = 0.40, P_{2,33} = 0.30. \end{aligned}$$

Case 7 (regime 1):

$$\begin{aligned} P_{1,11} &= 0.40, P_{1,12} = 0.30, P_{1,13} = 0.30, \\ P_{1,21} &= 0.20, P_{1,22} = 0.40, P_{1,23} = 0.40, \\ P_{1,31} &= 0.60, P_{1,32} = 0.20, P_{1,33} = 0.20. \end{aligned}$$

Case 7 (regime 2):

$$\begin{aligned} P_{2,11} &= 0.20, P_{2,12} = 0.40, P_{2,13} = 0.40, \\ P_{2,21} &= 0.20, P_{2,22} = 0.60, P_{2,23} = 0.60, \\ P_{2,31} &= 0.30, P_{2,13} = 0.40, P_{2,33} = 0.40. \end{aligned}$$

Case 7 (regime 3):

$$\begin{aligned} P_{3,11} &= 0.40, P_{3,12} = 0.30, P_{3,13} = 0.30, \\ P_{3,21} &= 0.20, P_{3,22} = 0.40, P_{3,23} = 0.40, \\ P_{3,31} &= 0.60, P_{3,32} = 0.20, P_{3,33} = 0.20. \end{aligned}$$

Case 8 (regime 1):

$$\begin{aligned} P_{1,11} &= 0.55, P_{1,12} = 0.25, P_{1,13} = 0.20, \\ P_{1,21} &= 0.25, P_{1,22} = 0.55, P_{1,23} = 0.20, \\ P_{1,31} &= 0.20, P_{1,32} = 0.25, P_{1,33} = 0.55. \end{aligned}$$

Case 8 (regime 2):

$$\begin{aligned} P_{2,11} &= 0.45, P_{2,12} = 0.20, P_{2,13} = 0.35, \\ P_{2,21} &= 0.30, P_{2,22} = 0.45, P_{2,23} = 0.25, \\ P_{2,31} &= 0.15, P_{2,32} = 0.40, P_{2,33} = 0.45. \end{aligned}$$

Case 8 (regime 3):

$$\begin{aligned} P_{3,11} &= 0.55, P_{3,12} = 0.25, P_{3,13} = 0.20, \\ P_{3,21} &= 0.25, P_{3,22} = 0.55, P_{3,23} = 0.20, \\ P_{3,31} &= 0.20, P_{3,32} = 0.25, P_{3,33} = 0.55. \end{aligned}$$

Case 9 (regime 1):

$$\begin{aligned} P_{1,11} &= 0.50, P_{1,12} = 0.30, P_{1,13} = 0.20, \\ P_{1,21} &= 0.50, P_{1,22} = 0.30, P_{1,32} = 0.20, \\ P_{1,31} &= 0.50, P_{1,32} = 0.30, P_{1,33} = 0.20. \end{aligned}$$

Case 9 (regime 2):

$$\begin{aligned}
P_{2,11} &= 0.30, P_{2,12} = 0.40, P_{2,13} = 0.30,\\
P_{2,21} &= 0.30, P_{2,22} = 0.40, P_{2,23} = 0.30,\\
P_{2,31} &= 0.30, P_{2,32} = 0.40, P_{2,33} = 0.30.
\end{aligned}$$

Case 9 (regime 3):

$$\begin{aligned}
P_{3,11} &= 0.50, P_{3,12} = 0.30, P_{3,13} = 0.20,\\
P_{3,21} &= 0.50, P_{3,22} = 0.30, P_{3,23} = 0.20,\\
P_{3,31} &= 0.50, P_{3,23} = 0.30, P_{3,33} = 0.20.
\end{aligned}$$

In general, the tests are consistent in the sense that the respective power tends to 1. The exception is Case 8, where the regimes exhibit some persistency and there are three different regimes, in that the $\sup LR(0 \mid 1)$ and the $WDmax(0 \mid 1, 2)$ presents low asymptotic power. Considering only two regimes, both the power of the $\sup LR(0 \mid 1)$ and of the $WDmax(0 \mid 1, 2)$ is satisfactory, even with high persistency.

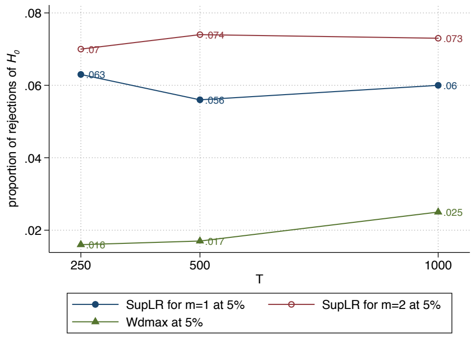
As expected, with one break (Cases 4, 5, 6) the $\sup LR(0 \mid 1)$ always performed better than the $WDmax(0 \mid 1)$. When two breaks are present, the $\sup LR(0 \mid 2)$ outperformed the $WDmax(0 \mid 1, 2)$, however the $WDmax(0 \mid 1, 2)$ surpassed the $\sup LR(0 \mid 1)$. This suggests that if we are not sure about the number of breaks the $WDmax$ might be a good option.

Furthermore, the $Seq(1 \mid 2)$ behaves as expected. It tends to correctly reject 1 break against 2 in all situations, both in small and in large samples, except for the case of highly persistent regimes (Figure 15b).

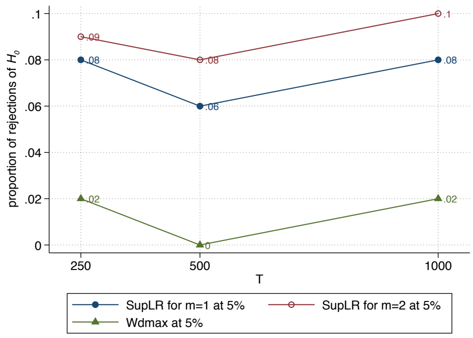
In small samples, with persistent regimes and in the presence of two breaks the tests tend to evidence lack of power, notably the $\sup LR(0 \mid 1)$ and the $WDmax(0 \mid 1, 2)$, Figure 16b. In turn, in finite samples with two breaks, against an alternative of one single break the null hypothesis of no break is only rejected about 18.3% of the time by the $\sup LR(0 \mid 2)$ and around 20% by the $WDmax(0 \mid 1, 2)$, which is far from being positive. However, the DGP was quite extreme, as the induced magnitude of the shifts is relatively small, the regimes are persistent in all of the three segments, and DGP of the first segment equals the third one.

When two breaks are present, the $\sup LR(0 \mid 2)$ always present a good power, larger than the $WDmax(0 \mid 1, 2)$ in all circumstances. This is the expected result, given that H_1 fully specified $\sup LR(0 \mid 2)$. In a nutshell, the higher the persistency of the regimes the lower the power of the tests. It must, however, be pointed out that the multiple breaks have been induced in such a way that they are difficult to detect: two breaks with the first and third regimes the same and small magnitudes [71, p 32].

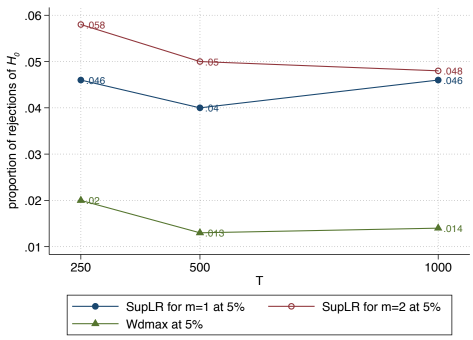
A practical recommendation of this exercise may involve the joint use of the various tests to detect and test inhomogeneities in a Markov chain. In fact, one can use the $\sup LR(0 \mid m)$ for several levels of m, in conjunction with a confirmatory analysis with the other tests. The $WDmax$ may be used with an arbitrary large M to confirm if there is at least one structural break, then the $Seq(\ell \mid \ell + 1)$ can be employed to corroborate the number of breaks of the process.



(a) Case 1

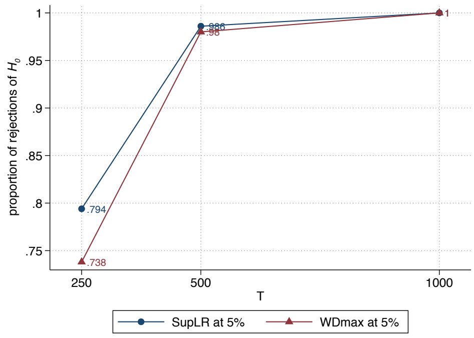


(b) Case 2

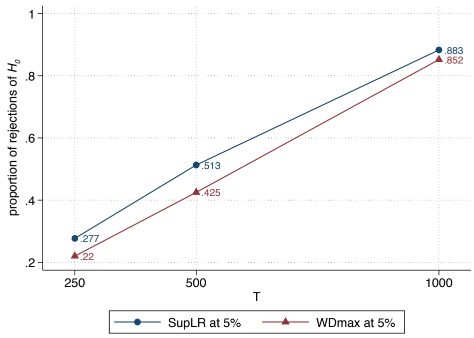


(c) Case 3

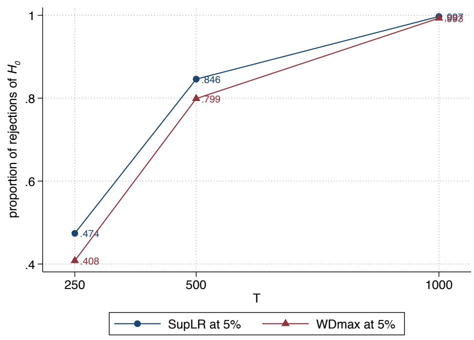
Figure 14: Homogeneous Markov chain, up to two breaks are allowed (nominal size: 5%)



(a) Case 4

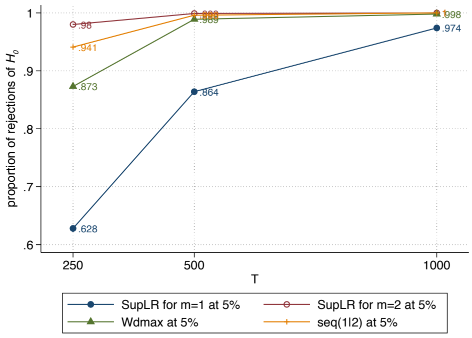


(b) Case 5

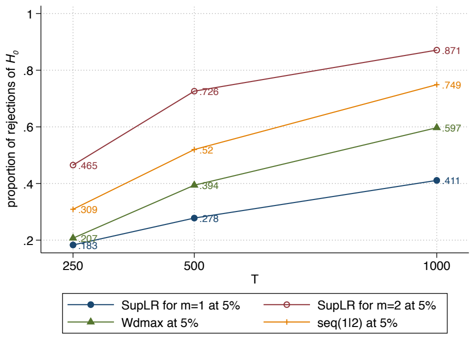


(c) Case 6

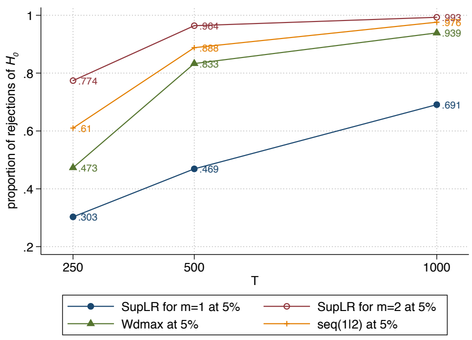
Figure 15: Inhomogeneous Markov chain with two segments, one break is allowed (nominal size: 5%)



(a) Case 7



(b) Case 8



(c) Case 9

Figure 16: Inhomogeneous Markov chain with three segments, two breaks are allowed (nominal size: 5%)

5.4 EXTENSIONS AND FURTHER RESEARCH

The research focus is to test for structural breaks in only some rows of the transition probability matrix, while keeping others unchanged. Such a test can be carried out by estimating the model subject to restrictions in the parameters. This ideia can be generalised for testing structural breaks in expected times, e.g. in the duration of bull and bear markets.

It could also be interesting to investigate the advantages in terms of forecasting of considering the inhomogeneous nature of a Markov chain. What are the practical consequences of ignoring the inhomogeneous nature of a Markov chain?

One natural extension to this work is to extend this analysis to multivariate Markov chains and to higher order Markov chains. Finally, this methodology is worthy to be applied empirically. For example, to analyse the predictability of stock returns testing the random walk hypothesis of stock prices; or to detect and test changing points in categorical time-series in general.

5.5 CONCLUSIONS

This article proposes a new approach for detecting and testing inhomogeneities in Markov chains occurring at unknown periods. The usual methods described in the literature for testing inhomogeneities in Markov chains have some limitations. Namely that they can only test for the presence of a single structural break; and either the break is assumed to occur at a known date or the limiting distribution of the test is unknown. Our strategy relies on the fact that, under certain conditions, an ergodic Markov chain admits a stable vectorial autoregressive representation.

The numerical equivalence between the MLE estimator for the one-step transition probabilities and the VAR mean parameter estimators is proved. Taking advantage of the possibility of representing a Markov chain in VAR form, the methods that usually apply to a VAR model remain valid for the Markov chains, namely, the sup LR, the $WDmax$, and the $Seq\,(\ell+1|\ell)$. These procedures are applied for the first time to Markov chains.

A Monte Carlo simulation study points to a higher power of sup LR tests compared to $WDmax$ tests when the alternative hypothesis is correctly specified. This evidence occurs with one and with two structural breaks.

As for size analysis, with one and with two breaks, there were no size distortions in either small or large samples, for all tests. With regard to power analysis, in general the tests were asymptotically consistent. Since the power increased substantially with T converging to 1. However, as DGP implied an increase in the persistence of regimes, the asymptotic power of the tests worsened slightly, in particular in the sup $LR(0 \mid 1)$ and in the $WDmax$ when two breaks are present.

Regarding further research we believe that inhomogeneity in Markov chains will continue to deserve analytical work as well as close care in accounting for its consequences in practical forecasting exercises.

APPENDIX 1: MATHEMATICAL APPENDIX

Propositon 5. *Under Assumptions A6 and A7, the OLS estimator for (51) is, regardless of the sample size, numerically equal to the one obtained through the ML that assumes a multinomial distribution in (49)*

Proof. The proof is immediate by the Frisch-Waugh-Lovell theorem, given that the variables of the right-hand side of the equation (45) are, by construction, orthogonal.

In fact, the stacked vector $\hat{\mathbf{p}}$ of the estimators $\hat{\mathbf{P}}_{\bullet i}$ is given by

$$\begin{aligned}
\hat{\mathbf{p}} &= \sum_{t=1}^{T} (\mathbf{x}_t \mathbf{x}_t')^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{y}_t \\
&= \begin{pmatrix} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' & 0 & \cdots & 0 \\ 0 & \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} \mathbf{z}_t y_{1t} \\ \sum_{t=1}^{T} \mathbf{z}_t y_{2t} \\ \vdots \\ \sum_{t=1}^{T} \mathbf{z}_t y_{q-1t} \end{pmatrix},
\end{aligned}$$

thus, the i-th estimator is given by

$$\hat{\mathbf{P}}_{\bullet i} = \left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \sum_{t=1}^{T} \mathbf{z}_t y_{it}.$$

In view of the fact that

$$\begin{aligned}
\mathbf{z}_t \mathbf{z}_t' &= \begin{pmatrix} y_{1t-1} \\ \vdots \\ y_{qt-1} \end{pmatrix} \begin{pmatrix} y_{1t-1} & \cdots & y_{qt-1} \end{pmatrix} \\
&= \begin{pmatrix} y_{1t-1}^2 & 0 & \cdots & 0 \\ 0 & y_{2t-1}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & y_{qt-1}^2 \end{pmatrix} = \begin{pmatrix} y_{1t-1} & 0 & \cdots & 0 \\ 0 & y_{2t-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & y_{qt-1} \end{pmatrix}
\end{aligned}$$

then

$$\left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \right)^{-1} = \begin{pmatrix} \frac{1}{\sum_{t=1}^{T} y_{1t-1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sum_{t=1}^{T} y_{2t-1}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{\sum_{t=1}^{T} y_{qt-1}} \end{pmatrix}.$$

Therefore, the k-th entry of $\hat{\mathbf{P}}_{\bullet i}$ is

$$\begin{aligned}\hat{P}_{ki} &= \frac{1}{\sum_{t=1}^{T} y_{kt-1}} \sum_{t=1}^{T} z_{kt} y_{it} \\ &= \frac{\sum_{t=1}^{T} y_{kt-1} y_{it}}{\sum_{t=1}^{T} y_{kt-1}} = \\ &= \frac{\sum 1\left(y_{kt-1} = 1, y_{it} = 1\right)}{\sum 1\left(y_{kt-1} = 1\right)} \\ &= \frac{\sum 1\left(S_{t-1} = k, S_t = i\right)}{\sum 1\left(S_{t-1} = k\right)} \\ &= \frac{n_{ik}}{n_i}.\end{aligned}$$

This expression numerically equals expression (49) as we have orthogonal partitioned regressions. ■

Propositon 6. *Under Assumptions A6 and A7 we have*

$$\ell_j^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} x_t x_t' \overset{a.s.}{\to} Q_j, \tag{59}$$

a non random positive definite matrix, as $\ell_j \to \infty$, *for each* $j = 1, \cdots, m+1$ *and* $\ell_j \leqslant T_j^0 - T_{j-1}^0 + 1$.

Proof. $\{S_t\}$ is a stationary and ergodic Markov chain sequence and $\mathbf{E}\left[|S_t|\right]$ is finite, therefore, using, for example, the pointwise ergodic theorem for stationary sequences [88, 98] we have

$$\begin{aligned}\frac{1}{\ell_j} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} z_t z_t' &= \left(\frac{1}{\ell_j} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} 1\left\{S_{t-1} = i\right\}\right) \\ &= \left(\frac{1}{\ell_j} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} y_{it-1}\right)_{i=1,\cdots,q} \\ &\overset{as}{\to} E_j\left[y_{it}\right] = \pi_i(j) \end{aligned} \tag{88}$$

Using the continuous mapping theorem

$$\frac{1}{\ell_j} \sum \mathbf{x}_t \mathbf{x}_t{}' \overset{as}{\to} \mathbf{\Pi}_j,$$

where $\mathbf{\Pi}_j$ is the vector of stationary probabilities for the j-th segment. ■

Propositon 7. *Under Assumptions A6 and A7:*

1. $\{\mathbf{x}_t \varepsilon_t, \mathcal{F}_t\}$ *forms a martingale difference sequence;*
2. $\mathbf{E}\left[\mathbf{x}_t \varepsilon_t\right] = 0$;

Proof. The vector $\boldsymbol{\varepsilon}_t$ is, by construction, a martingale difference sequence with respect to $\mathcal{F}_t$, given that $\varepsilon_{it} = y_{it} - \mathbf{E}[y_{it} \mid \mathcal{F}_t]$. Hence, $\{\mathbf{x}_t \boldsymbol{\varepsilon}_t\}$ is also a martingale difference sequence because $\mathbf{E}[\mathbf{x}_t \boldsymbol{\varepsilon}_t \mid \mathcal{F}_t] = \mathbf{x}_t \mathbf{E}[\boldsymbol{\varepsilon}_t \mid \mathcal{F}_t] = 0$. Therefore, $\mathbf{E}[\mathbf{x}_t \boldsymbol{\varepsilon}_t] = 0$ (by the law of iterated expectations), $\{\mathbf{z}_t \boldsymbol{\varepsilon}_t\}$ is an uncorrelated sequence and as a direct consequence $\{\mathbf{x}_t \boldsymbol{\varepsilon}_t, \mathcal{F}_t\}$ forms a martingale difference sequence and, thus, a strongly mixing sequence.

■

Lemma 1. *The covariance matrix (55)* $\boldsymbol{\Sigma} \equiv \mathbf{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t']$ *given by*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \pi_1 - \sum_{k=1}^q \pi_k P_{k1}^2 & -\sum_{k=1}^q P_{k1} P_{k2} \pi_k & \cdots & -\sum_{k=1}^q P_{k1} P_{kq} \pi_k \\ -\sum_{k=1}^q P_{k1} P_{k2} \pi_k & \pi_2 - \sum_{k=1}^q \pi_k P_{k2}^2 & \cdots & -\sum_{k=1}^q P_{k2} P_{kq} \pi_k \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{k=1}^q P_{k1} P_{kq} \pi_k & -\sum_{k=1}^q P_{k2} P_{kq} \pi_k & & \pi_q - \sum_{k=1}^q \pi_k P_{kq}^2 \end{pmatrix}.$$

Proof. The covariance writes:

$$\begin{aligned} \mathbf{E}\left[\varepsilon_{it}\varepsilon_{ij}\right] &= \mathbf{E}\left[\left(y_{it} - x_t' P_{\bullet i}\right)\left(y_{jt} - x_t' P_{\bullet j}\right)\right] = \\ &= \mathbf{E}\left[y_{it} y_{jt} - y_{it} x_t' P_{\bullet j} - x_t' P_{\bullet i} y_{jt} + x_t' P_{\bullet i} x_t' P_{\bullet j}\right] \\ &= \mathbf{E}\left[y_{it} y_{jt}\right] - \mathbf{E}\left[y_{it} x_t' P_{\bullet j}\right] - \mathbf{E}\left[x_t' P_{\bullet i} y_{jt}\right] + \mathbf{E}\left[x_t' P_{\bullet i} x_t' P_{\bullet j}\right] \\ &= 0 - \sum_{k=1}^q P_{kj} P_{ki} \pi_k - \sum_{k=1}^q P_{ki} P_{kj} \pi_k + \sum_{k=1}^q \pi_k P_{ki} P_{kj} \\ &= -\sum_{k=1}^q P_{ki} P_{kj} \pi_k. \end{aligned}$$

Since, by construction $y_{it} y_{jt} = 0, i \neq j$, then

$$\begin{aligned} \mathbf{E}\left[y_{it} y_{j,t-1}\right] &= P\left(S_t = i, S_{t-1} = j\right) = P\left(S_t = i \middle| S_{t-1} = j\right) P\left(S_{t-1} = j\right) \\ &= P_{ji} \pi_j \\ \mathbf{E}\left[y_{it} x_t' P_{\bullet j}\right] &= \mathbf{E}\left[y_{it}\left(y_{1,t-1} P_{1j} + \ldots + y_{q,t-1} P_{qj}\right)\right] = \sum_{k=1}^q P_{kj} \mathbf{E}\left[y_{it} y_{k,t-1}\right] \\ &= \sum_{k=1}^q P_{kj} P_{ki} \pi_k \\ \mathbf{E}\left[x_t' P_{\bullet i} y_{jt}\right] &= \mathbf{E}\left[y_{jt} x_t' P_{\bullet i}\right] = \sum_{k=1}^q P_{ki} P_{kj} \pi_k \\ \mathbf{E}\left[x_t' P_{\bullet i} x_t' P_{\bullet j}\right] &= \mathbf{E}\left[P_{\bullet i}' x_t x_t' P_{\bullet j}\right] = P_{\bullet i}' \mathbf{E}\left[x_t x_t'\right] P_{\bullet j} = \\ &= P_{\bullet i}' \begin{bmatrix} \pi_1 & 0 & \cdots & 0 \\ 0 & \pi_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_q \end{bmatrix} P_{\bullet j} = \sum_{k=1}^q \pi_k P_{ki} P_{kj}. \end{aligned}$$

On the other hand, $\mathbf{E}\left[\varepsilon_{it}^2\right]$ may be written as

$$\begin{aligned}\mathbf{E}\left[\varepsilon_{it}^2\right] &= \mathbf{E}\left[\left(y_{it} - x_t' P_{\bullet i}\right)^2\right] = \\ &= \mathbf{E}\left[y_{it}^2 - 2y_{it}x_t'P_{\bullet i} + \left(x_t'P_{\bullet i}\right)^2\right] \\ &= \mathbf{E}\left[y_{it}^2 - 2y_{it}x_t'P_{\bullet i} + P_{\bullet i}'x_t x_t' P_{\bullet i}\right]\end{aligned}$$

and

$$\begin{aligned}\mathbf{E}\left[y_{it}y_{it}\right] &= \mathbf{E}\left[y_{it}^2\right] = \mathbf{E}\left[y_{it}\right] = \pi_i \\ \mathbf{E}\left[y_{it}x_t'P_{\bullet i}\right] &= \sum_{k=1}^{q} P_{kj}^2 \pi_k \\ \mathbf{E}\left[P_{\bullet i}'x_t x_t' P_{\bullet i}\right] &= \sum_{k=1}^{q} \pi_k P_{ki}^2,\end{aligned}$$

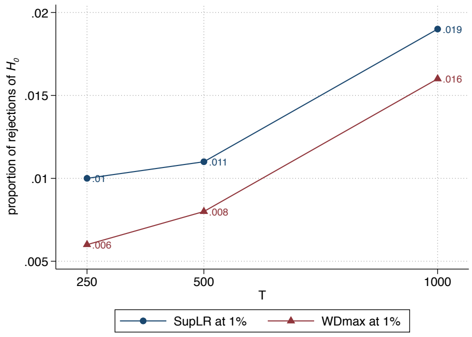
therefore

$$\begin{aligned}\mathbf{E}\left[\varepsilon_{it}^2\right] &= \pi_i - 2\sum_{k=1}^{q} P_{kj}^2 \pi_k + \sum_{k=1}^{q} \pi_k P_{ki}^2 \\ &= \pi_i - \sum_{k=1}^{q} \pi_k P_{ki}^2.\end{aligned}$$

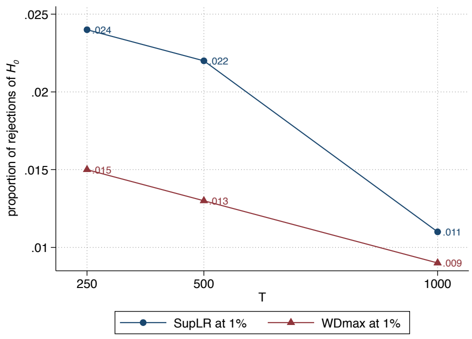
■

APPENDIX 2: MONTE CARLO SIMULATION RESULTS

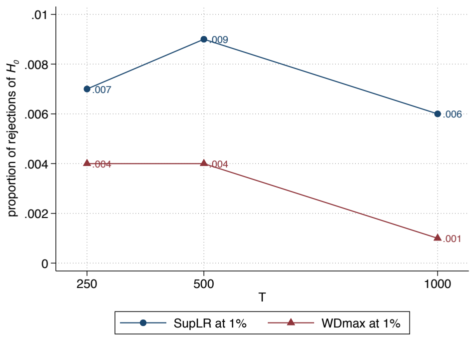
Size Analysis



(a) Case 1

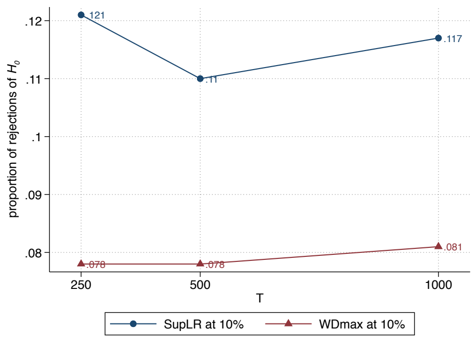


(b) Case 2

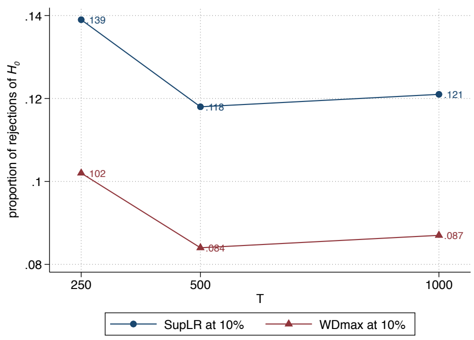


(c) Case 3

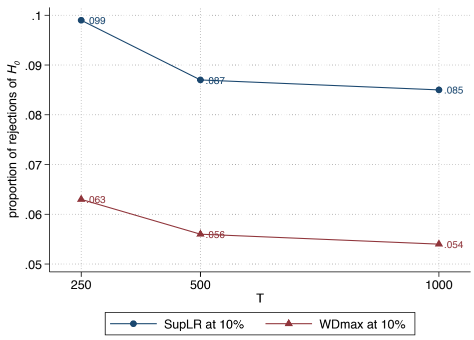
Figure 17: Homogeneous Markov chain, one break is allowed (nominal size: 1%)



(a) Case 1

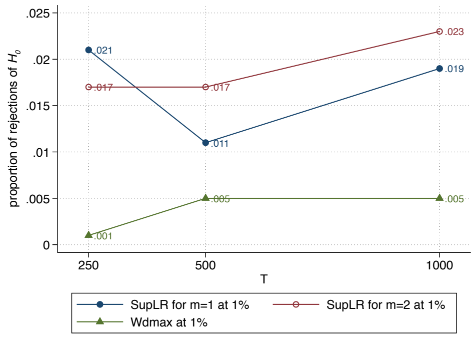


(b) Case 2

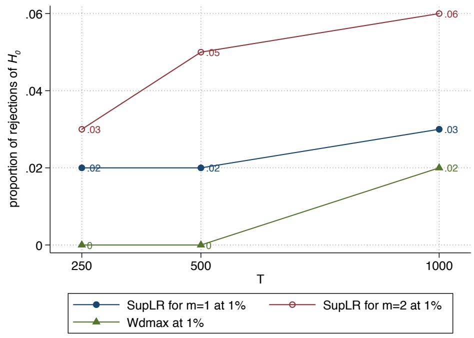


(c) Case 3

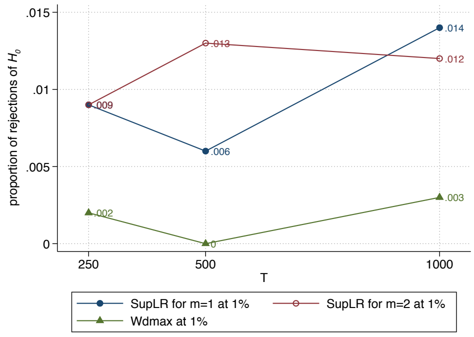
Figure 18: Homogeneous Markov chain, one break is allowed (nominal size: 10%)



(a) Case 1

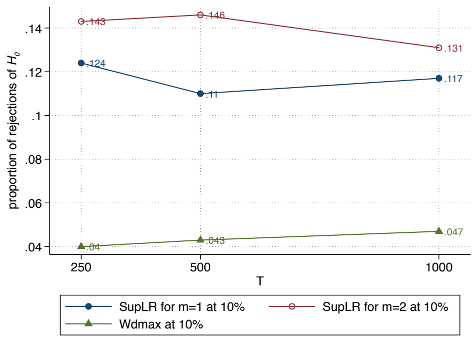


(b) Case 2

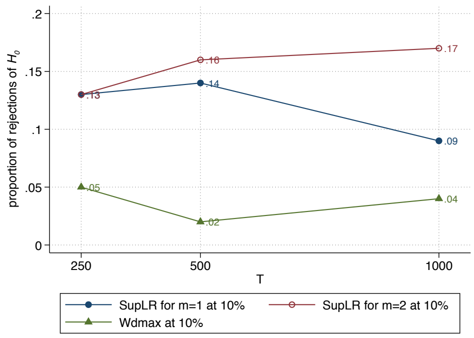


(c) Case 3

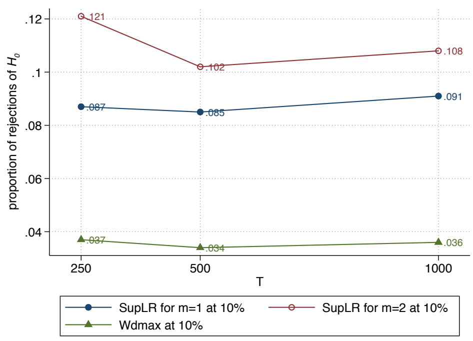
Figure 19: Homogeneous Markov chain, up to two breaks are allowed (nominal size: 1%)



(a) Case 1



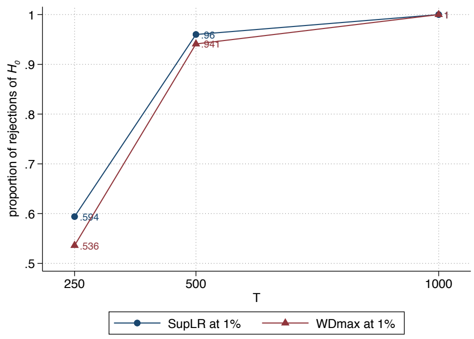
(b) Case 2



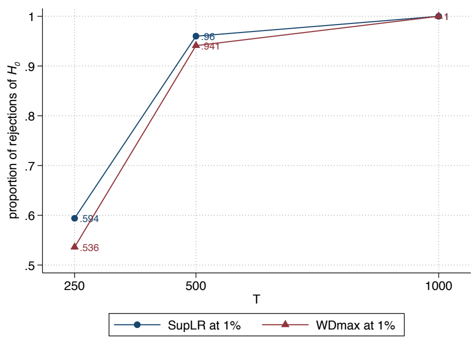
(c) Case 3

Figure 20: Homogeneous Markov chain, up to two breaks are allowed (nominal size: 10%)

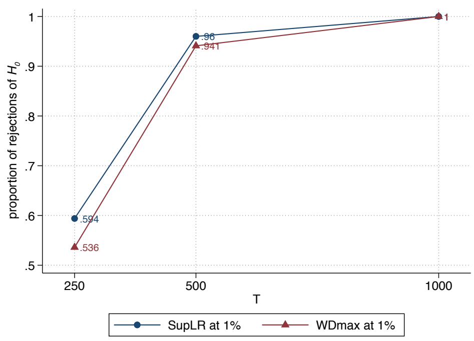
Power Analysis



(a) Case 4

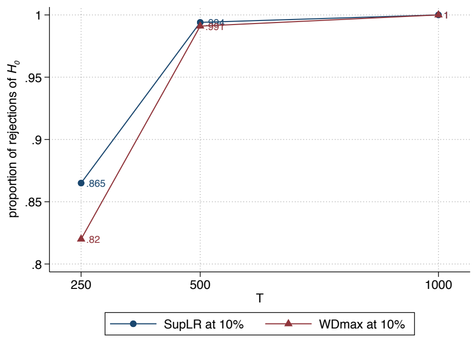


(b) Case 5

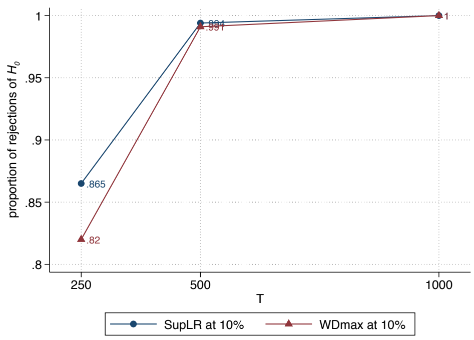


(c) Case 6

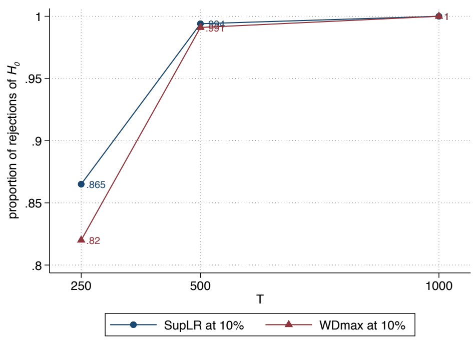
Figure 21: Inhomogeneous Markov chain with two segments, one break is allowed (nominal size: 1%)



(a) Case 4

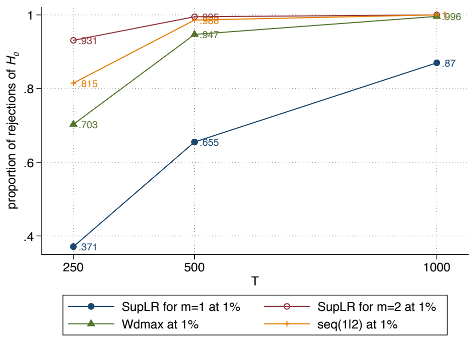


(b) Case 5

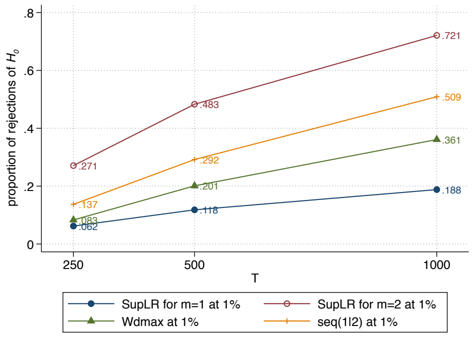


(c) Case 6

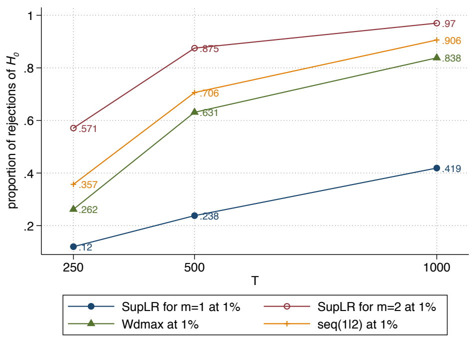
Figure 22: Inhomogeneous Markov chain with two segments, one break is allowed (nominal size: 10%)



(a) Case 7

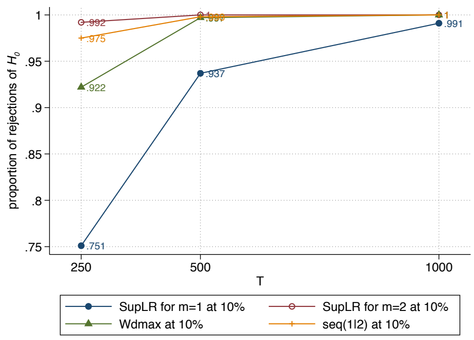


(b) Case 8

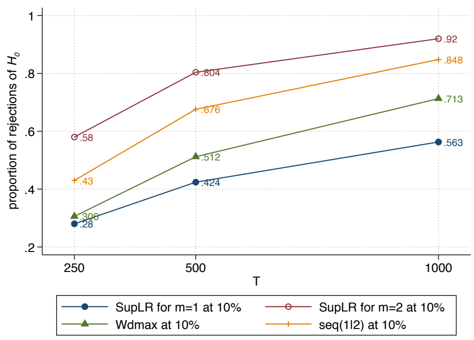


(c) Case 9

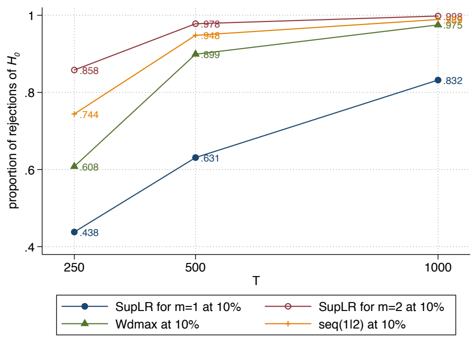
Figure 23: Inhomogeneous Markov chain with three segments, two breaks are allowed (nominal size: 1%)



(a) Case 7



(b) Case 8



(c) Case 9

Figure 24: Inhomogeneous Markov chain with three segments, two breaks are allowed (nominal size: 10%)

6 CONCLUDING REMARKS

This Thesis contributes to the Markov chains literature by developing methods focused on the estimation issue that are usually circumscribed to the continuous variables. The first essay, *Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition*, considered the multivariate Markov chain model applied to an example from economic history, more precisely to the creative destruction in ocean-going shipping technologies during the early XIX century: steam and sailing. In fact, the struggle between sailing and steam is a long-standing theme in economic history, but this technological competition story has only partly tackled. Moreover, we compare a classical multivariate linear econometric approach (a VAR model) with a multivariate Markov chain. While the former fails to detect linear interdependency relationships between these two technologies, the latter method detects evidence that the relationship was nonlinear, with a strong indication of complementarities and cross-technology learning effects. To the best of our knowledge, this has been the first application of Markov chains to an instance of economic history. It suggested that Markov chains are a valuable tool and fertile instrument that can be used in an interdisciplinary manner, even when applied in non-standard realities.

The second essay, *Combining a regression model with a multivariate Markov chain in a forecasting problem*, addressed the situation where a multivariate Markov chain, with an arbitrary large domain and number of series, play the role of covariates in a regression model. In a situation where a dependent variable depends on categorical variables, the past state interactions between latter can be modelled through a multivariate Markov chain to more accurately forecast the future values of the former. In fact, exploiting the intra and inter-transition probabilities within and between data categories may lead to a substantial forecasting improvement of that endogenous variable. To the best of our knowledge this has never been done before. This concept was illustrated through a Monte Carlo simulation that pointed out the benefits of the proposed method over some alternative methods.

The third essay, *The changing economic regimes and expected time to recover of the peripheral countries under the euro: a nonparametric approach*, proposes a simple non-parametric model to estimate the expected time that a stochastic process takes to cross a defined threshold based on a Markov chain representation of such a process. An economic application suggested that the Euro generated a regime change in the macrodynamics of Europe and that this change impacted on the growth of the economies involved, triggering a process of divergence. These two essays illustrate the flexibility of the Markov chain approach and how it may be deployed in situations that go beyond the simple calculation of transition probabilities. Put another way, it shows that the calculated transition probabilities can be useful to different nontrivial diligences, for example, to assist the forecast of continuous variables or to compute expected hitting times. An interesting extension is to consider the issue of inhomogeneity either for the computation of the expected hitting times and for the concept of inhomogeneous Markov chains as covariates. Furthermore, the specification considered in Chapter 3 was linear. It could also be meaningful to incorporate nonlinearities in the functional form that governs the relationship between the dependent variable and the multivariate Markov chain.

Finally, in Chapter 5, *Time inhomogeneous multivariate Markov chains: detecting and testing multiple structural breaks occurring at unknown dates*, we propose a methodology for testing multiple structural breaks occurring at unknown date intervals in multivariate Markov chains. Even though several studies approached Markov chains, few studies focused on the issue of nonhomogeneity. More importantly, the issue of detecting and testing multiple distinct transition probability matrices in a Markov chain has not been studied yet. Our approach relies on a vectorial autoregressive representation of the Markov chain. The method presented here can be used to test for a large number of breaks in a Markov chain, as it showed great flexibility. Our proposal can also be extended to multivariate Markov chains and to higher order Markov chains, provided the sample size is large enough. On top of that, to overcome the curse of dimensionality problems, it might be interesting to develop a method to test the presence of structural breaks in the framework of the MTD-Probit. These topics are currently under research. There are grounds to argue that the literature in Markov chains might assume the probable systematic inhomogeneous nature of the underlying stochastic process.

BIBLIOGRAPHY

[1] Robert C Allen. *Global economic history: a very short introduction*. Vol. 282. Oxford University Press, 2011.

[2] João Ferreira do Amaral. *Porque devemos sair do Euro*. Leya, 2013.

[3] Giovanni Arrighi. *The long 20th century. Money, power, and the origins of our times*. 1995.

[4] Mehdi Rajabi Asadabadi. "A customer based supplier selection process that combines quality function deployment, the analytic network process and a Markov chain." In: *European Journal of Operational Research* 263.3 (2017), pp. 1049–1062.

[5] KB Athreya and Cheng-Der Fuh. "Bootstrapping Markov chains: countable case." In: *Journal of Statistical Planning and Inference* 33.3 (1992), pp. 311–331.

[6] Jushan Bai. "Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance-Covariance Matrices." In: *Annals of Economics and Finance* 1.2 (2000), pp. 303–339.

[7] Jushan Bai, Robin L Lumsdaine, and James H Stock. "Testing for and dating common breaks in multivariate time series." In: *The Review of Economic Studies* 65.3 (1998), pp. 395–432.

[8] Jushan Bai and Pierre Perron. "Estimating and testing linear models with multiple structural changes." In: *Econometrica* (1998), pp. 47–78.

[9] Jushan Bai and Pierre Perron. "Computation and analysis of multiple structural change models." In: *Journal of applied econometrics* 18.1 (2003), pp. 1–22.

[10] Jushan Bai and Pierre Perron. "Multiple structural change models: a simulation analysis." In: *Econometric theory and practice: Frontiers of analysis and applied research* (2006), pp. 212–237.

[11] Ishwar V Basawa. *Statistical Inferences for Stochasic Processes: Theory and Methods*. Elsevier, 2014.

[12] Andre Berchtold. "Estimation in the mixture transition distribution model." In: *Journal of Time Series Analysis* 22.4 (2001), pp. 379–397.

[13] Patrick Billingsley. "Statistical methods in Markov chains." In: *The Annals of Mathematical Statistics* (1961), pp. 12–40.

[14] Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I Del Genio, Jesús Gómez-Gardenes, Miguel Romance, Irene Sendina-Nadal, Zhen Wang, and Massimiliano Zanin. "The structure and dynamics of multilayer networks." In: *Physics Reports* 544.1 (2014), pp. 1–122.

[15] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity." In: *Journal of econometrics* 31.3 (1986), pp. 307–327.

[16] Samuel Bowles, David M Gordon, and Thomas E Weisskopf. "Power and profits: the social structure of accumulation and the profitability of the postwar US economy." In: *Review of Radical Political Economics* 18.1-2 (1986), pp. 132–167.

[17] Robert Boyer and Yves Saillard. *Régulation Theory: the state of the art*. Routledge, 2005.

[18] Bruce Bukiet, Elliotte Rusty Harold, and José Luis Palacios. "A Markov chain approach to baseball." In: *Operations Research* 45.1 (1997), pp. 14–23.

[19] Julia García Cabello. "The future of branch cash holdings management is here: New Markov chains." In: *European Journal of Operational Research* 259.2 (2017), pp. 789–799.

[20] Ding-Geng Chen and YL Lio. "A novel estimation approach for mixture transition distribution model in high-order Markov chains." In: *Communications in Statistics-Simulation and Computation* 38.5 (2009), pp. 990–1003.

[21] Wai-Ki Ching, Eric S Fung, and Michael K Ng. "A multivariate Markov chain model for categorical data sequences and its applications in demand predictions." In: *IMA Journal of Management Mathematics* 13.3 (2002), pp. 187–199.

[22] Wai-Ki Ching and Michael K Ng. "Markov chains." In: *Models, algorithms and applications* (2006).

[23] Robin Craig. *British tramp shipping, 1750-1914 / Robin Craig*. English. International Maritime Economic History Association St. John's, 2004, vi, 408 p. : ISBN: 0973007346.

[24] Bruno Damásio, Francisco Louçã, and João Nicolau. "The changing economic regimes and expected time to recover of the peripheral countries under the euro: A nonparametric approach." In: *Physica A: Statistical Mechanics and its Applications* 507 (2018), pp. 524–533.

[25] Bruno Damásio and Sandro Mendonça. "Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition." In: *Applied Economics Letters* (2018), pp. 1–7.

[26] Bruno Damásio and João Nicolau. "Combining a regression model with a multivariate Markov chain in a forecasting problem." In: *Statistics & Probability Letters* 90 (2013), pp. 108–113.

[27] James Davidson. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford, 1994.

[28] Francis X Diebold and Robert S Mariano. "Comparing predictive accuracy." In: *Journal of Business & economic statistics* 20.1 (2002).

[29] Barry Eichengreen. *Hall of Mirrors: The Great Depression, the Great Recession, and the Uses-and Misuses-of History*. Oxford University Press, 2014.

[30] Robert F Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." In: *Econometrica: Journal of the Econometric Society* (1982), pp. 987–1007.

[31] Robert F Engle, David F Hendry, and Jean-Francois Richard. "Exogeneity." In: *Econometrica: Journal of the Econometric Society* (1983), pp. 277–304.

[32] Larrie D Ferreiro and Alexander Pollara. "Contested Waterlines: The Wave-Line Theory and Shipbuilding in the Nineteenth Century." In: *Technology and culture* 57.2 (2016), pp. 414–444.

[33] Chris Freeman and Francisco Louçã. *As time goes by: the information revolution and the industrial revolutions in historical perspective*. Oxford University Press, 2001.

[34] Eric S Fung and Tak Kuen Siu. "A flexible Markov chain approach for multivariate credit ratings." In: *Computational Economics* 39.2 (2012), pp. 135–143.

[35] Frank W Geels. "Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study." In: *Research policy* 31.8-9 (2002), pp. 1257–1274.

[36] Sergio Gómez, Alexandre Arenas, J Borge-Holthoefer, Sandro Meloni, and Yamir Moreno. "Discrete-time Markov chain approach to contact-based disease spreading in complex networks." In: *EPL (Europhysics Letters)* 89.3 (2010), p. 38009.

[37] Adam Gottschau. "Exchangeability in multivariate Markov chain models." In: *Biometrics* (1992), pp. 751–763.

[38] Steven Gray. *Channelling mobilities: migration and globalisation in the Suez Canal region and beyond, 1869–1914*. 2015.

[39] James D Hamilton. "A new approach to the economic analysis of nonstationary time series and the business cycle." In: *Econometrica: Journal of the Econometric Society* (1989), pp. 357–384.

[40] James D Hamilton. "Specification testing in Markov-switching time-series models." In: *Journal of econometrics* 70.1 (1996), pp. 127–157.

[41] Bruce E Hansen. "The new econometrics of structural change: dating breaks in US labour productivity." In: *Journal of Economic perspectives* 15.4 (2001), pp. 117–128.

[42] Peter Reinhard Hansen. "Structural changes in the cointegrated vector autoregressive model." In: *Journal of Econometrics* 114.2 (2003), pp. 261–295.

[43] Knick Harley. "The shift from sailing ships to steamships, 1850-1890: A study in technological change and its diffusion." In: *Essays on a Mature Economy: Britain After 1840*. Ed. by D. N. McCloskey. London: Methuen, 1971, pp. 215–237.

[44] Philip A Horvath, Chad W Autry, and William E Wilcox. "Liquidity implications of reverse logistics for retailers: A Markov chain approach." In: *Journal of retailing* 81.3 (2005), pp. 191–203.

[45] Morgan Kelly and Cormac O'Grada. "Speed under Sail during the Early Industrial Revolution." In: (2018).

[46] Paul Krugman. "The return of depression economics and the crisis of 2008." In: *New York* (2009).

[47] Sophie Lèbre and Pierre-Yves Bourguignon. "An EM algorithm for estimation in the mixture transition distribution model." In: *Journal of Statistical Computation and Simulation* 78.8 (2008), pp. 713–729.

[48] Y Li, M Zhu, R Klein, and N Kong. "Using a partially observable Markov chain model to assess colonoscopy screening strategies–A cohort study." In: *European Journal of Operational Research* 238.1 (2014), pp. 313–326.

[49] Alain Lipietz and Ian Patterson. *The enchanted world: Inflation, credit and the world crisis*. Verso London, 1985.

[50] Richard G Lipsey, Kenneth I Carlaw, and Clifford T Bekar. *Economic transformations: general purpose technologies and long-term economic growth*. Oxford University Press, 2005.

[51] Nuno Luís Madureira. "Oil in the age of steam." In: *Journal of Global History* 5.1 (2010), pp. 75–94.

[52] Andrei A Markov. "Primer statisticheskogo issledovanija nad tekstomEvgenija Onegina'illjustrirujuschij svjaz'ispytanij v tsep (An example of statistical study on the text ofEugene Onegin'illustrating the linking of events to a chain)." In: *Izvestija Imp. Akad. nauk* (1913).

[53] Jun-ichi Maskawa. "Multivariate Markov chain modeling for stock markets." In: *Physica A: Statistical Mechanics and its Applications* 324.1-2 (2003), pp. 317–322.

[54] Ed McKenzie. "Some simple models for discrete variate time series 1." In: *JAWRA Journal of the American Water Resources Association* 21.4 (1985), pp. 645–650.

[55] Farhad Mehran. "Longitudinal analysis of employment and unemployment based on matched rotation samples." In: *Labour* 3.1 (1989), pp. 3–20.

[56] Sandro Mendonça. "The sailing ship effect: Reassessing history as a source of insight on technical change." In: *Research Policy* 42.10 (2013), pp. 1724–1738.

[57] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[58] BR Mitchell. *British Historical Statistics*. Cambridge: Cambridge University Press, 1988.

[59] Wesley Clair Mitchell. *Business cycles and their causes*. Univ of California Press, 1959.

[60] Saif I Shah Mohammed and Jeffrey G Williamson. "Freight rates and productivity gains in British tramp shipping 1869–1950." In: *Explorations in Economic History* 41.2 (2004), pp. 172–203.

[61] Christian J Murray and David H Papell. "The purchasing power parity persistence paradigm." In: *Journal of International Economics* 56.1 (2002), pp. 1–19.

[62] Whitney K Newey and Daniel McFadden. "Large sample estimation and hypothesis testing." In: *Handbook of econometrics* 4 (1994), pp. 2111–2245.

[63] João Nicolau. *Modelação de Séries Temporais Financeiras*. Coleção Económicas - 2ª Série, Almedina, 2012.

[64] João Nicolau. "A new model for multivariate Markov chains." In: *Scandinavian Journal of Statistics* 41.4 (2014), pp. 1124–1135.

[65] João Nicolau. "A simple nonparametric method to estimate the expected time to cross a threshold." In: *Statistics & Probability Letters* (2016).

[66] João Nicolau. "A simple nonparametric method to estimate the expected time to cross a threshold." In: *Statistics & Probability Letters* 123 (2017), pp. 146–152.

[67] João Nicolau and Flavio Ivo Riedlinger. "Estimation and inference in multivariate Markov chains." In: *Statistical Papers* 56.4 (2015), pp. 1163–1173.

[68] Douglass North. "Ocean freight rates and economic development 1730-1913." In: *The Journal of Economic History* 18.4 (1958), pp. 537–555.

[69] Luigi Pascali. "The wind of change: Maritime technology, trade, and economic development." In: *American Economic Review* 107.9 (2017), pp. 2821–54.

[70] Pierre Perron. "The great crash, the oil price shock, and the unit root hypothesis." In: *Econometrica: Journal of the Econometric Society* (1989), pp. 1361–1401.

[71] Pierre Perron. "Dealing with structural breaks." In: *Palgrave handbook of econometrics* 1.2 (2006), pp. 278–352.

[72] Alan M Polansky. "Detecting change-points in Markov chains." In: *Computational Statistics & Data Analysis* 51.12 (2007), pp. 6013–6026.

[73] Sidney Pollard and Paul Robertson. *The British Shipbuilding Industry: 1870-1914*. Harvard University Press, 1979.

[74] Zhongjun Qu and Pierre Perron. "Estimating and testing structural changes in multivariate regressions." In: *Econometrica* 75.2 (2007), pp. 459–502.

[75] Adrian E Raftery. "A model for high-order Markov chains." In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1985), pp. 528–539.

[76] Adrian Raftery and Simon Tavaré. "Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution model." In: *Applied Statistics* (1994), pp. 179–199.

[77] Peter Rappoport and Lucrezia Reichlin. "Segmented trends and non-stationary time series." In: *The Economic Journal* 99.395 (1989), pp. 168–177.

[78] Nathan Rosenberg. "Factors affecting the diffusion of technology." In: *Explorations in economic history* 10.1 (1972), pp. 3–33.

[79] Ahmet D Sahin and Zekai Sen. "First-order Markov chain approach to wind speed modelling." In: *Journal of Wind Engineering and Industrial Aerodynamics* 89.3-4 (2001), pp. 263–269.

[80] Joseph Alois Schumpeter. *Business cycles*. Vol. 1. McGraw-Hill New York, 1939.

[81] Bruno Sericola. *Markov chains: theory and applications*. John Wiley & Sons, 2013.

[82] A Shamshad, MA Bawadi, WMA Wan Hussin, TA Majid, and SAM Sanusi. "First and second order Markov chain models for synthetic generation of wind speed time series." In: *Energy* 30.5 (2005), pp. 693–708.

[83] Christopher A Sims. "Macroeconomics and reality." In: *Econometrica: Journal of the Econometric Society* (1980), pp. 1–48.

[84] Tak-Kuen Siu, Wai-Ki Ching, S Eric Fung, and Michael K Ng. "On a multivariate Markov chain model for credit risk measurement." In: *Quantitative Finance* 5.6 (2005), pp. 543–556.

[85] Peter M Solar. "Opening to the East: Shipping between Europe and Asia, 1770–1830." In: *The Journal of Economic History* 73.3 (2013), pp. 625–661.

[86] Joseph E Stiglitz. *The price of inequality: How today's divided society endangers our future*. WW Norton & Company, 2012.

[87] Joseph E Stiglitz. *The euro: How a common currency threatens the future of Europe*. WW Norton & Company, 2016.

[88] William F Stout. *Almost sure convergence*. Tech. rep. 1974.

[89] Y Suhov and M Kelbert. *Markov Chains: A Primer in Random Processes and their Applications, Cambrigde*. 2008.

[90] Lawrence H Summers. "The age of secular stagnation: what it is and what to do about it." In: *Foreign Aff.* 95 (2016), p. 2.

[91] Barış Tan and Kamil Yılmaz. "Markov chain test for time dependence and homogeneity: an analytical and empirical evaluation." In: *European Journal of Operational Research* 137.3 (2002), pp. 524–543.

[92] George Tauchen. "Finite state markov-chain approximations to univariate and vector autoregressions." In: *Economics letters* 20.2 (1986), pp. 177–181.

[93] Howell Tong. *Threshold models in non-linear time series analysis*. New York: Springer Lecture Notes in Statistics, 1983.

[94] Ruey S Tsay. "Testing and modeling threshold autoregressive processes." In: *Journal of the American statistical association* 84.405 (1989), pp. 231–240.

[95] Christos Aristeides Tsiliyannis. "Markov chain modeling and forecasting of product returns in remanufacturing based on stock mean-age." In: *European Journal of Operational Research* (2018).

[96] Peter B Turchin. "Modelling the effect of host patch size on Mexican bean beetle emigration." In: *Ecology* 67.1 (1986), pp. 124–132.

[97] Immanuel Wallerstein. *The capitalist world-economy*. Vol. 2. Cambridge University Press, 1979.

[98] Halbert White. *Asymptotic theory for econometricians*. Academic press, 2014.

[99] Dong-Mei Zhu and Wai-Ki Ching. "A new estimation method for multivariate Markov chain model with application in demand predictions." In: *Business Intelligence and Financial Engineering (BIFE), 2010 Third International Conference on*. IEEE. 2010, pp. 126–130.